Scleroderma is an autoimmune disease characterized by microvascular injury and varying degrees of widespread fibrosis of both skin and internal organs, and can be subtyped depending on particular patterns of disease [1].

There is increasing evidence suggesting genetic involvement in the pathogenesis of scleroderma. Several genes are thought to be involved in an epistatic fashion in combination with environmental stimuli. Studies have reported the association of scleroderma with particular MHC class II alleles, cytokine gene polymorphisms, and associations with genes involved in extracellular matrix (ECM) turnover and oxidative stress [2].

A major cause of death in scleroderma patients is pulmonary dysfunction as a result of progressive lung disease. The pathology of pulmonary fibrosis displays features of abnormal wound repair with excessive fibroblast proliferation and deposition of ECM, leading to progressive fibrosis and loss of lung function. Genetic factors are also likely to be involved in determining the pattern of organ involvement.

SPARC (secreted protein, acidic and rich in cysteine), also known as osteonectin or BM-40, is a matricellular protein found in the ECM with expression restricted to tissues undergoing consistent turnover as well as to sites of injury and disease [3]. SPARC binds a number of ECM proteins and growth factors. Along with fibronectin and collagen, SPARC is a component of the ECM and mediates cell–cell and cell–matrix interactions, and is thought to have anti-adhesive properties that can assist cell migration [4–6]. Immunohistochemical analysis of open lung biopsy material from patients with usual interstitial pneumonia revealed the localization of SPARC to fibroblastic foci [7]. In other fibrotic conditions, such as hepatic fibrosis and kidney fibrosis, SPARC is localized to sites of diseased tissue [8, 9].

SPARC mRNA was found to be increased in fibroblasts cultured from the affected skin of scleroderma patients [10]. Whole-genome expression profiling confirmed this finding that SPARC mRNA is increased in fibroblasts isolated from both affected and non-affected skin in scleroderma patients compared with controls [11]. In a subsequent study, circulating SPARC was shown to be increased in scleroderma patients with limited disease compared with those with diffuse disease and controls [12].

Clinical Genomics Group, National Heart and Lung Institute and Royal Brompton Hospital, Imperial College, Pan-pathology Molecular Diagnostics Group, St Thomas' Hospital and Centre for Rheumatology, Royal Free Hospital and University College Medical School, London, UK.

Submitted 26 July 2004; revised version accepted 30 September 2004.

Correspondence to: K. I. Welsh, Clinical Genomics Group, National Heart and Lung Institute and Royal Brompton Hospital, Imperial College, 1B Manresa Road, London SW3 6LR, UK. E-mail: k.welsh@imperial.ac.uk

S. M. du Bois and K. I. Welsh

Objective. SPARC (secreted protein, acidic and rich in cysteine) is a matricellular protein that modulates cell–cell and cell–extracellular matrix interactions. SPARC expression is restricted mainly to sites of tissue remodelling and wound repair, and is prominent in fibrotic disorders. Single-nucleotide polymorphisms (SNPs) in the SPARC gene are reportedly linked to scleroderma in four ethnic groups: Choctaw Indians, Caucasians, African Americans and Mexican Americans. We set out to reproduce and to positionally clone these disease associations in a set of UK Caucasian scleroderma patients and ethnically matched controls.

Methods. One hundred and twenty-one scleroderma subjects and 200 controls were genotyped by polymerase chain reaction with sequence-specific primers differing only in the 3′ nucleotide corresponding to each allele of the biallelic SNPs. Scleroderma patients were analysed against controls and on the basis of their fibrosing alveolitis status as judged by high-resolution computed tomography evaluation and the extent of cutaneous involvement.

Results. Eight biallelic SNPs were genotyped: three from the last untranslated exon, which had been described previously, and an additional five novel SNPs: two in the promoter region, one in exon three and two in the 3′ untranslated region. Six major haplotypes were constructed across all eight SNP positions. No significant differences in genotype, allele or haplotype frequency were observed between scleroderma and controls or within scleroderma subgroups.

Conclusions. SNPs in the SPARC gene are not associated with susceptibility to scleroderma. This research adds to the genetic knowledge of the SPARC gene by identifying five novel SNPs spanning the whole gene and inserting these within the context of clearly defined haplotypes.

Key words: SPARC, Scleroderma, Fibrosis, SNP, Haplotype.
Three single-nucleotide polymorphisms (SNPs) in the *SPARC* gene at positions 998, 1551 and 1922 (numbered erroneously as 1992), all residing in the 3' untranslated region (UTR) of the gene, have been identified by Zhou *et al.* [13]. Within this cohort of patients, homozygosity for the 998C allele was found to be increased in scleroderma patients across ethnic lines, whereas heterozygosity at this locus was decreased. In addition, subgroup analysis identified the 1551G allele and the 1922T allele to be over-represented in patients with Raynaud’s phenomenon and pulmonary fibrosis, respectively. The authors also discovered that homozygosity for the 998C 3' UTR allele was associated with increased mRNA stability.

From the NCBI SNP database, we recently identified the presence of five additional SNPs: two in the promoter region, one in exon 3 and two in the 3' UTR. We examined the distribution of these novel SNPs in addition to the three SNPs examined by Zhou *et al.* in our set of 121 UK Caucasian patients and 200 healthy controls. Furthermore, we constructed the single-nucleotide variant haplotypes defined by these eight SNPs. Genotyping of all SNPs except 1922 was as follows. From each sample, primers polymerase chain reaction method (SSP-PCR) method. To ascertain the genotyping accuracy of the SSP method, two of the polymorphisms, +12647 (998) C/G and +13200 (1551) G/C, were sequenced in 25 individuals. We set out to reproduce and map further these disease associations in a UK Caucasian scleroderma set while expanding these single SNP alleles into multiple SNP haplotypes.

**Materials and methods**

**Patients and controls**

One hundred and twenty-one UK Caucasian scleroderma patients were recruited for this study, of which 99 were female and 22 were male. Written consent was obtained from all patients and the study was approved by Ethics Committees at the Royal Brompton and Royal Free Hospitals. All patients fulfilled the American College of Rheumatology preliminary criteria for scleroderma [1]. When classified according to disease extent, 97 patients were classified as having limited cutaneous disease and 24 patients as having diffuse cutaneous disease. Fibrosing alveolitis was present in 61 patients and absent in 60, as judged by high-resolution computed tomography (HRCT) evaluation. The control group consisted of 200 healthy UK Caucasian subjects.

**SSP-PCR conditions**

Genomic DNA from patients and controls was extracted from peripheral blood. Genotyping was performed using sequence-specific primers that differ only in the 3' nucleotide corresponding to each allele of the biallelic SNPs [14]. Details of primer mixes are available at *Rheumatology* Online as supplementary data. Genotyping of all SNPs except 1922 was as follows. From each primer mix, 5 µl aliquots were dispensed into each well of a 96-well plate (Costar, The Netherlands). The other reagents were mixed with an overlay of 10 µl mineral oil (Bioline). PCR amplification was performed on an MJ Research 96V or PTC200 machine (GRI) using the following cycling parameters: 2 min at 96°C followed by 26 cycles of 96°C for 30 s, 56°C for 45 s and 72°C for 1 min, with a final step of 2 min at 72°C.

Genotyping accuracy was assessed by means of DNA sequencing to confirm SSP-PCR results for SNPs +12647 (998) C/G and +13200 (1551) G/C. Sequencing was performed on a CEQ8000 capillary sequencer (Beckman Coulter, High Wycombe, UK) according to the manufacturer's instructions on PCR products generated using the primer sequences 5'-CTGCAATGTGTTGTGT TAAAGGC-3' and 5'-GGGAAATGCGTTATCATCGA-3' for SNP 998 and 5'-GACAATCATCAATGAGG-3' and 5'ACAGACACACGATGTTGCTAAG-3' for SNP 1551. PCR products were as described above for SNP 1922 with the substitution of Promega *Taq* DNA polymerase in place of that from Bioline and a melting temperature of 58°C. The same primers were used for the sequencing reaction. The Qiagist kit (Qiagen, Crawley, UK) was used for PCR product purification according to the manufacturer's instructions.

**DNA sequencing**

Genotyping and allele frequencies were calculated by direct counting. Haplotypes were constructed by eye and subsequently checked by the haplotype construction program Arlequin version 2.0 ([http://anthropologie.unige.ch/arlequin/](http://anthropologie.unige.ch/arlequin/)). Data were initially analysed by the data mining program Knowledge Seeker (Angoss Software, Guildford, UK), and subsequently by χ² contingency tables or Fisher's exact test when appropriate. Power calculations were performed with the PS power and sample size calculation program by Dupont and Plummer ([http://www.mc.vanderbilt.edu/prevmed/ps/index.htm](http://www.mc.vanderbilt.edu/prevmed/ps/index.htm)).

**Results**

The genotype frequencies for each SNP conformed to Hardy-Weinberg equilibrium. Genotype frequencies are displayed in Table 1 and SNPs are numbered with reference to their position from the start codon, as obtained from http://snpper.chip.org/bio/snpper-enter-gene. Haplotype construction produced six clear haplotypes across the eight SNPs covering 93% of the fully
C/G and agreement 100% of the time for SNP by SSP-PCR. Results obtained from the two methods were in 25 samples for SNP same results were obtained. These individuals were genotyped a second time and the pairwise haplotype between SNPs 12647 and 13571 was not as expected. These individuals were genotyped a second time and the same results were obtained. Twenty-five individuals were sequenced for SNPs +12647 (998), +13200 (1551) and +13571 (1922). *Genotyping available for 198 controls and 101 SSc patients only.

### Table 1. Genotype frequencies of SPARC SNPs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Control</th>
<th>Total SSc</th>
<th>FASSc</th>
<th>NFASSc</th>
<th>LeSSc</th>
<th>DcSSc</th>
</tr>
</thead>
<tbody>
<tr>
<td>−11663 AA</td>
<td>60 (0.30)</td>
<td>31 (0.26)</td>
<td>14 (0.23)</td>
<td>17 (0.28)</td>
<td>27 (0.28)</td>
<td>4 (0.17)</td>
</tr>
<tr>
<td>−11663 AG</td>
<td>90 (0.45)</td>
<td>61 (0.50)</td>
<td>33 (0.54)</td>
<td>28 (0.47)</td>
<td>47 (0.48)</td>
<td>14 (0.58)</td>
</tr>
<tr>
<td>−11663 GG</td>
<td>50 (0.25)</td>
<td>29 (0.24)</td>
<td>14 (0.23)</td>
<td>15 (0.25)</td>
<td>23 (0.24)</td>
<td>6 (0.25)</td>
</tr>
<tr>
<td>−962 GG</td>
<td>67 (0.33)</td>
<td>36 (0.28)</td>
<td>21 (0.34)</td>
<td>15 (0.25)</td>
<td>28 (0.29)</td>
<td>8 (0.33)</td>
</tr>
<tr>
<td>−962 GA</td>
<td>93 (0.47)</td>
<td>59 (0.49)</td>
<td>29 (0.48)</td>
<td>30 (0.50)</td>
<td>50 (0.51)</td>
<td>9 (0.38)</td>
</tr>
<tr>
<td>−962 AA</td>
<td>40 (0.20)</td>
<td>26 (0.21)</td>
<td>11 (0.18)</td>
<td>15 (0.25)</td>
<td>19 (0.20)</td>
<td>7 (0.29)</td>
</tr>
</tbody>
</table>

### Table 2. SPARC haplotypes

<table>
<thead>
<tr>
<th>Haplotype frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>−11663</td>
</tr>
<tr>
<td>SSc</td>
</tr>
<tr>
<td>Haplotype 1</td>
</tr>
<tr>
<td>Haplotype 2</td>
</tr>
<tr>
<td>Haplotype 3</td>
</tr>
<tr>
<td>Haplotype 4</td>
</tr>
<tr>
<td>Haplotype 5</td>
</tr>
<tr>
<td>Haplotype 6</td>
</tr>
</tbody>
</table>

Absolute numbers stated with frequencies in parenthesis (%). Total SSc, scleroderma patients as a whole; FASSc, scleroderma patients with fibrosing alveolitis; NFASSc, scleroderma patients without fibrosing alveolitis; LeSSc, scleroderma patients with limited cutaneous disease; DcSSc, scleroderma patients with diffuse cutaneous disease. There was no statistically significant difference in genotype frequency between groups. Relative positions from the start codon were obtained from http://snpper.chip.org/bio/snpper-enter-gene. Numbering corresponding to Zhou et al. [13] is as follows; +12647 (998), +13200 (1551) and +13571 (1922).

No significant differences in genotype, allele or haplotype frequency were found between the Caucasian scleroderma patients and controls (Tables 1 and 2). Scleroderma patients were subgrouped according to the presence or absence of fibrosing alveolitis, as judged by HRCT evaluation and limited or diffuse cutaneous disease extent. We found no significant differences between patients within these groups.

**Discussion**

SPARC is thought to be a pivotal protein in regulating cell-cell and cell-matrix interactions via its counteradhesive function. In addition, it can modify extracellular matrix deposition, influence angiogenesis and alter the activity of a number of cytokines and...
growth factors. Thus, genetic variation in the SPARC gene, especially in the promoter and 3' UTR region, could well be involved in the regulation of mRNA production and stability, and consequently may predispose individuals to excessive fibrogenesis. This makes SPARC a good candidate gene for genetic association studies in scleroderma. However, we found no significant associations between the eight SPARC SNPs or their haplotypes with susceptibility to scleroderma, the presence of fibrosing alveolitis or the extent of cutaneous involvement in our set of Caucasian subjects.

In this study, we obtained results conflicting with those previously published. Zhou et al. reported SPARC SNP 998 CC homozygotes to be significantly increased and CG heterozygotes significantly decreased in scleroderma patients across ethnic lines compared with controls [17]. We did not reproduce these associations in our Caucasian cohort. Although we cannot completely rule out the possibility of an association, we have confidence in our negative findings. Power calculations based on the 998 CC association (odds ratio 1.99, 95% confidence interval 0.95, 3.78) in the Caucasian subjects from Zhou's paper indicate that we have 85% power to detect the same odds ratio with 121 patients and 200 controls. This does not, however, rule out the possibility of a lesser degree of difference between scleroderma and controls. One explanation for the 998 association in the Zhou paper is that the control genotype frequencies have deviated from Hardy–Weinberg equilibrium. In addition, due to the heterogeneity of the disease in question, we cannot disregard the possibility of population stratification giving rise to discordant results between the two studies.

In our study, we also did not observe the previously reported association between SNP 1922 and pulmonary fibrosis found by Zhou et al. [13]. Moreover, nearly 100% linkage disequilibrium was evident between SNPs 998 and 1922 in both our patient and control groups. In this regard, our findings of no disease association with either the 998 polymorphism or the 1922 polymorphism are quite consistent with our observation of the tight linkage between these two polymorphisms.

Zhou et al. also reported a genome-wide association identifying, among others, a microsatellite marker close to the SPARC gene as linked with scleroderma in Choctaw Indians [13]. The results of our study are based only on Caucasian subjects and hence do not exclude the possibility that there is a genetic link between SPARC and scleroderma in different ethnic groups.

Increased SPARC expression in scleroderma patients cannot be disputed as more than one group has reported this finding. However, it is possible that the up-regulation of SPARC expression in these patients could be the result of the increased expression of transforming growth factor β (TGF-β) and other related factors, known to occur in scleroderma patients [16]. TGF-β induces SPARC expression in fibroblasts, keratinocytes, smooth muscle, endothelial cells and epithelial cells; and indeed the reverse is also true as SPARC can induce TGF-β expression in mesangial and epithelial cells. It is widely accepted that TGF-β is a profibrotic growth factor capable of inducing the production of extracellular matrix proteins. In scleroderma, there are conflicting reports as to whether the putative expression of TGF-β is increased in fibrotic lesions as a whole. However, in a study by Higley et al., TGF-β1 was localized within scleroderma skin biopsies using immunocytochemistry staining, and increased TGF-β1 staining was reported in patients with limited disease compared with patients with diffuse disease [17]. These data complement the findings of Macko et al., who found that patients with limited disease had increased circulating concentrations of SPARC compared with patients with diffuse disease and controls [12].

We observed no significant association between patterns of inherited single-nucleotide variations across the SPARC gene and Caucasian scleroderma patients both as a whole and within the context of lung fibrosis and cutaneous involvement. Therefore, the raised SPARC levels in these patients do not appear to have their origins in a genetically predisposing component within the SPARC gene. This indicates that the increases in mRNA expression and circulating protein are likely to be in response to profibrotic factors such as TGF-β. This research adds to our genetic knowledge of the SPARC gene by identifying additional SNPs spanning the whole gene and inserting these within the context of clearly defined haplotypes.

<table>
<thead>
<tr>
<th>Key messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• SNPs in the SPARC gene are not associated with susceptibility to scleroderma.</td>
</tr>
<tr>
<td>• Eight SNPs, including five novel SNPs spanning the SPARC gene, can be constructed into six clear haplotypes.</td>
</tr>
</tbody>
</table>

**Acknowledgements**

This work was supported by the Raynaud’s and Scleroderma Association and the Arthritis Research Campaign, UK. The authors wish to thank Dr Paolo Spagnolo (Clinical Genomics Group, National Heart and Lung Institute and Royal Brompton Hospital, Imperial College, London) for his assistance in the collection and evaluation of clinical data, and Dr David Abraham (Centre for Rheumatology, Royal Free Hospital and University College Medical School, London) for his contribution to the preparation of the manuscript.

The authors have declared no conflicts of interest.

**Supplementary data**

Supplementary data are available at *Rheumatology Online.*

**References**


