Genetics and Genomic Studies in Scleroderma (Systemic Sclerosis)

Sandeep K. Agarwal, MD, PhD*, Filemon K. Tan, MD, PhD, Frank C. Arnett, MD

Division of Rheumatology, Department of Internal Medicine, The University of Texas Health Science Center at Houston, 6431 Fannin, MSB 5.270, Houston, TX 77030, USA

Scleroderma or systemic sclerosis is an autoimmune connective tissue disease clinically characterized by fibrosis of the skin and internal organs and obliteratorive vasculopathy. The complexities of scleroderma are evident from the variability in its clinical manifestations, which probably reflects the diverse mechanisms that underlie the development of disease subtypes. Despite recent advances in the understanding of some of the molecular pathways involved in scleroderma, the etiopathogenesis remains unknown. Although fibrosis and endothelial dysfunction are hallmarks of the disease, autoimmunity is probably the root cause. Autoimmunity and inflammation currently are best exemplified by the multiple but not overlapping patterns of specific autoantibodies in patients who have scleroderma. In fact, each of these autoantibodies tends to mark a distinct clinical subset of disease [1]. The presence of inflammatory infiltrates in the dermis early in the disease and increased circulating levels of cytokines and chemokines in patients who have scleroderma further implicate inflammation in the pathogenesis of scleroderma. It remains unknown how these autoimmune responses lead to certain patterns of organ damage that vary among different clinical subsets of scleroderma.

It currently is believed that scleroderma is a complex polygenic disease that occurs in genetically predisposed individuals who have encountered specific environment exposures and/or other stochastic factors. The nature of these genetic determinants and how they interact with environmental factors are areas of active investigation. This article discusses the evidence that
supports a strong genetic link to scleroderma. Also reviewed are the family and twin studies that suggest a genetic component in scleroderma, and recent genetic-association studies implicating specific genes in the pathoge-netic triad of autoimmunity, endothelial dysfunction, and fibroblast activation. Last, this article highlights recent studies in scleroderma that use gene-expression microarray profiles seeking to identify pathogenetic pathways in scleroderma. Together these studies implicate potential pathogenetic mechanisms involved in scleroderma, which, it is hoped, may translate into clinical utility, including determination of disease risk, diagnosis, prognosis, and novel therapeutics.

**Familial aggregation and twin studies**

Determining that a disease occurs more commonly in families than in the general population is a first step in implicating potential genetic contributions. Robust estimates of prevalence and incidence rates of scleroderma, which are essential in determining familial aggregation, have varied widely in the general population, ranging from 3.1 to 20.8 per 100,000 and from 0.4 to 1.2 per 100,000, respectively [2–5]. A recent, well-conducted study estimated the prevalence of scleroderma to be 24.2 per 100,000 adults with an annual incidence of 1.93 per 100,000 adults [6]. Initial reports suggested that familial clustering was uncommon, but only recently has it been quantitated [7]. Ten cases of scleroderma in first-degree relatives among 710 proband cases were reported in the Sydney, Australia population, conferring a familial risk conservatively estimated at 11 (95% confidence interval [CI], 2.7–19.3) [8]. It was demonstrated subsequently that scleroderma recurred in 1.6% of families of scleroderma cases in three separate cohorts that had an estimated population risk of only 0.026% [9]. Although this absolute risk of familial scleroderma was relatively low, the familial relative risk was approximately 15-fold higher for siblings and 13-fold higher for first-degree relatives. A recent study suggested that the affected members within multicase scleroderma families tend to have concordant scleroderma-specific autoantibodies, further supporting the concept of a genetic predisposition [10]. Based on these studies, a positive family history of scleroderma confers the strongest known relative risk for disease.

The investigation of monozygotic twins is an important approach for assessing and quantifying the role of genetic versus environmental factors in specific diseases. Case reports have described concordance for scleroderma in twins [11,12]. Examination of 42 monozygotic and dizygotic twins collected from across the United States demonstrated similar scleroderma concordance rates (∼5%), thus implying no genetic susceptibility [13]. The concordance rate of antinuclear antibodies was significantly higher in monozygotic twins (90%) than in dizygotic twins (40%), however [13]. A subsequent study compared gene-expression microarray profiles of cultured fibroblasts from 15 discordant mono- and dizygotic twin pairs [14].
Unsupervised hierarchical clustering segregated cultured fibroblasts into two distinct groups. Fibroblast lines from unaffected monozygotic, but not dizygotic, twins tended to group with cultured fibroblast cell lines from affected scleroderma patients rather than with normal controls. Together, as summarized in Table 1, these data suggest that genetic factors may play a significant role in susceptibility to scleroderma with regards to the production of autoantibodies and in vitro fibroblast activation. Although these data may partially explain the clustering of autoimmune diseases observed in some families, it is necessary to define the genetic factors that underlie scleroderma susceptibility.

Ethnic factors

It is apparent that ethnicity influences the susceptibility to autoimmune diseases, including scleroderma. African Americans have been reported to have a higher incidence of scleroderma (22.5 cases per million per year) than white women (12.8 cases per million per year) [15]. Furthermore, African American women have been reported likely to have more severe disease, earlier age of onset, and worse survival rates [15]. Reveille and colleagues [16] demonstrated in a prospective cohort that Hispanics and African Americans were more likely than whites to have diffuse skin involvement, digital ulcerations, and pulmonary hypertension.

Among the more intriguing observations with regards to ethnicity and scleroderma were those made in the Oklahoma Choctaw Indians [17]. The prevalence in full-blooded Choctaws was estimated at 469 cases per 100,000 over a 4-year period. This prevalence was significantly higher than that seen in non–full-blooded Choctaws, other Native Americans in Oklahoma, and whites. Furthermore, a majority of the Choctaw scleroderma cases had diffuse scleroderma, pulmonary fibrosis, and circulating anti-topoisomerase I antibodies and could be traced genealogically to a common founding family in the late 1700s. No environmental factors were identified in the Oklahoma environment, and the strongest risk factor was a nearly unique Amerindian HLA class II haplotype (DRB1*1602, DQA1*0501, DQB1*0301). Investigators using microsatellite markers in

Table 1
Estimates of scleroderma occurrence and risk in families

<table>
<thead>
<tr>
<th>Population</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population (prevalence)</td>
<td>0.026</td>
</tr>
<tr>
<td>First-degree relatives of persons who have scleroderma (prevalence)</td>
<td>1.60</td>
</tr>
<tr>
<td>Siblings of persons who have scleroderma (prevalence)</td>
<td>0.40</td>
</tr>
<tr>
<td>Monozygotic twins of persons who have scleroderma (concordance)</td>
<td>5.00</td>
</tr>
<tr>
<td>Monozygotic twins with antinuclear antibody–positive concordance</td>
<td>90</td>
</tr>
<tr>
<td>Monozygotic twins with fibroblast gene profile concordance</td>
<td>50</td>
</tr>
</tbody>
</table>
three candidate regions found a shared haplotype on chromosome 15q containing the fibrillin-1 (FBN1) gene, an important structural protein and regulator of transforming growth factor-beta (TGF-β) within the extracellular matrix, to be significantly overrepresented in Choctaw scleroderma cases [18]. Similar studies of single-nucleotide polymorphisms (SNPs) supported a role for FBN1 in Choctaw and Japanese scleroderma cases but did not show associations in whites [19]. Interestingly, a mouse model of scleroderma, the tight skin mouse (tsk-1), is caused by a genomic duplication of fibrillin-1 [20].

These studies demonstrate the influence of ethnicity on scleroderma susceptibility. Although multiple factors including socioeconomic factors, access to health care, and even environmental exposures may help explain these difference in scleroderma susceptibility or even clinical expression, genetic differences among ethnic groups probably are key determinants as well.

**Genetic factors**

**HLA associations**

The major histocompatibility complex (MHC) or HLA region is the most polymorphic region of the genome. Polymorphisms in HLA have been linked to a number of autoimmune diseases including rheumatoid arthritis, ankylosing spondylitis, systemic lupus erythematosus, and many others [1,21–23]. Scleroderma also has been associated with HLA polymorphisms, and these associations have been reviewed previously [24,25]. The associations of HLA polymorphisms with scleroderma susceptibility itself are modest but, more importantly, are consistent and are reproducible across different populations; these include associations with the HLA-DR5/11 and DR3 haplotypes in white patients and with HLA-DR2 haplotypes in Japanese and Choctaw Indian patients. Of particular interest is a finding of a significantly higher frequency of \( HLA-DQA1^*0501 \) in male patients who have scleroderma [26].

Stronger associations with HLA haplotypes exist for specific autoantibody subsets of scleroderma, including HLA \( DRB1^*1104 \) and, independently, \( DPB1^*1301 \) in whites, \( DQB1^*0301 \) and \( DPB1^*1301 \) in African Americans, and DR2 haplotypes in Japanese (\( DRB1^*1502, DQB1^*0601, DPB1^*090 \)) and Choctaws (\( DRB1^*1602, DQB1^*0301, DPB1^*1301 \)) with anti-topoisomerase antibody [27–32]. Furthermore, HLA \( DQBJ^*0501 \) and other DQB1 alleles encoding nonpolar amino acids in position 26 are associated with anti-centromere antibody [27,33]. An association of HLA \( DRB1^*1302, DQB1^*0604/0605 \) haplotypes has been found with anti-fibrillarlin (anti-U3-RNP)–positive patients, who are more often male African Americans, and the HLA \( DRB1^*0301 \) haplotype has been shown to be associated with anti-PM-Scl antibody positivity in patients who tend to be nearly exclusively white [34,35]. Finally, an association was observed with
HLA \texttt{DQB1*0201} in patients who have anti-RNA polymerase I, II, and/or III antibodies, but this association has not been observed in other studies [32,36,37].

**Non-HLA candidate-gene associations**

Genetic-association studies seek to determine genetic variants associated with disease states or specific traits. As more studies have been undertaken in different complex diseases, it has become clear that the contribution of individual genes to the genetic risk for disease may be quite modest (relative risk, $\sim 1.5$–2.0) and that multiple loci are involved. In this light, interpretation of genetic-association studies in an uncommon and phenotypically heterogeneous disease, such as scleroderma, must be performed using strict guidelines. Such studies often are limited by lack of sufficient statistical power to generate reliable and reproducible results because of small sample sizes in cases and controls, genetic heterogeneity, and the extent and degree of linkage disequilibrium among genetic markers that vary among populations [38,39]. Population stratification, differences in phenotype of complex diseases, and quality control also complicate the interpretation of genetic-association studies. The inability to replicate data across study cohorts, however, may represent real biologic differences in populations that originate from different genetic backgrounds, as has been shown recently in rheumatoid arthritis [40–45]. Nonetheless, it is necessary to replicate the findings of association studies in additional appropriately designed cohorts. Furthermore, candidate genes identified in genetic-association studies must be followed by functional studies, usually done in vitro, that demonstrate that these associations are, in fact, causal or at least biologically plausible.

Choosing which targets to investigate is based largely on current paradigms of pathogenesis from human and/or animal studies or, in some instances, on associations with other autoimmune diseases. Thus, selection of these candidate genes is colored by publication bias and incomplete scientific knowledge. The candidate genes for scleroderma that have been investigated have been related largely to autoimmunity/inflammation, vascular function, and fibrosis or extracellular matrix production.

**Autoimmunity/Inflammation**

**Interleukin-1 alpha and beta**

Interleukin-1 alpha (IL-1$\alpha$) and interleukin-1 beta (IL-1$\beta$) are proinflammatory cytokines involved in a number of autoimmune diseases. Patients who have scleroderma have increased circulating levels of IL-1$\alpha$ and IL-1$\beta$, and IL-1$\alpha$ is a potent stimulatory factor of cultured dermal scleroderma fibroblast behavior in vitro [46–48]. In a study comparing 86 Japanese patients who had scleroderma with 70 healthy controls, the CTG haplotype
at positions −889, +4729, and +4845 of the \textit{IL1A} gene was associated with scleroderma susceptibility as well as the presence of interstitial lung disease [49]. Attempts to confirm this association have not been successful. The −889C allele of \textit{IL1A}, which is part of this haplotype, was not found to be associated with scleroderma in cohorts of Slovak and Italian patients [50,51]. In contradiction to the initial observation, the −889T allele of \textit{IL1A} was found to be associated with scleroderma in Slovak patients [50]. These studies must be interpreted cautiously because of the small sample sizes. Alternatively, the conflicting results may be caused by differences in linkage disequilibrium within the \textit{IL1} gene cluster among these populations.

Genetic associations with \textit{IL1B} also have been investigated in patients who have scleroderma. Mattuzzi and colleagues [52] recently demonstrated significant associations of the \textit{IL1B}-31-C and \textit{IL1B}-511-T alleles with susceptibility to scleroderma. Given the potential importance of IL-1β in scleroderma, it will be interesting to see if these genetic associations are replicated. Furthermore, additional genes involved in the IL-1α and IL-1β pathways may be involved in scleroderma susceptibility and should be considered for future studies.

\textit{Allograft inflammatory factor-1}

Allograft inflammatory factor-1 (AIF-1) is a newly identified protein identified in rat cardiac allografts undergoing chronic rejection [53]. The function and regulation of AIF-1 expression has not been characterized thoroughly. AIF-1 is expressed by macrophages and neutrophils, and immunohistochemical analyses of scleroderma biopsies have demonstrated increased AIF-1 expression in the infiltrating cells within lesions [54–56]. A recent study demonstrated that AIF-1 promoted T-cell infiltration and induced the expression types I and III collagen and cytokines by normal dermal fibroblasts, providing additional evidence for a role of AIF-1 in scleroderma [57]. A comparison of 140 patients who had scleroderma with 97 controls demonstrated an association of the \textit{AIF1} +889A allele with scleroderma [58]. This polymorphism, which is located in exon 3 of AIF-1, generates a nonsynonymous change (tryptophan to arginine). Investigators using the relatively large cohort of 548 patients who had scleroderma in the Scleroderma Family Registry and DNA Repository and an additional 467 patients from the Genetics versus Environment in Scleroderma Outcomes Study noted a modest association with this AIF-1 polymorphism, but only with the anti-centromere antibody (ACA)–positive subset of scleroderma [59]. Although the \textit{AIF1} gene maps within the MHC class II region, the association of \textit{AIF1} with ACA-positive scleroderma was not explained completely by linkage disequilibrium, suggesting an independent association of \textit{AIF1} with ACA-positive scleroderma [59]. Additional studies are needed to confirm these associations and to advance the understanding of the functional importance of AIF-1 in the pathogenesis of scleroderma.
Protein tyrosine phosphatase non-receptor 22

Protein tyrosine phosphatase non-receptor 22 (PTPN22) is a class I cysteine-based tyrosine phosphatase expressed in hematopoietic cells. In functional studies, PTPN22 exerts a negative regulatory function of T-cell receptor signaling [60,61]. The PTPN22 R620W mutation has been shown to increase this negative regulatory effect on T-cell receptor signaling [62]. The PTPN22 R620W mutation has been associated with several diseases, including type I diabetes mellitus, systemic lupus erythematosus, and rheumatoid arthritis, but not ankylosing spondylitis, Crohn’s disease, or primary Sjögren’s syndrome [43,45,63–67]. Two initial reports with small sample sizes failed to show an association of PTPN22 and scleroderma [68,69]. This finding is not surprising, given the modest genetic risk conferred by PTPN22 for diseases such as rheumatoid arthritis (odds ratio [OR], ~1.5) [43,45]. In contrast, a study combining three different cohorts, totaling 1120 patients who had scleroderma and 816 control subjects, found an association between the PTPN22 R620W SNP and both anti-topoisomerase-I (OR, 2.14; 95% CI, 1.4–3.2) and ACA-positive scleroderma (OR, 1.67; 95% CI, 1.2–2.4) [70]. This large study cohort provides strong evidence that inherent immunoregulatory defects common to some, but not all, autoimmune diseases also may play a role in scleroderma.

Macrophage migration inhibitory factor

Macrophage migration inhibitory factor (MIF) is a proinflammatory cytokine that is released from activated T cells and macrophages [71,72]. MIF promotes production of tumor necrosis factor-alpha and other cytokines by macrophages and T cells [71,72]. MIF expression is increased in patients who have rheumatoid arthritis and recently has been shown to be increased in scleroderma as well [73,74]. MIF polymorphisms were characterized in a cohort of 486 patients who had scleroderma and 254 healthy subjects [75]. The −173C MIF promoter polymorphism and the associated haplotype containing −173C and −794 7-CATT (C7) were found to be underrepresented in patients who had limited scleroderma compared with patients who had diffuse scleroderma or healthy controls. The genetic-association study was followed up by complementary functional studies of the effects of the polymorphisms that emphasize its potential relevance in disease pathogenesis. Fibroblasts from patients with this haplotype produced increased levels of MIF in vitro, consistent with a functional role of the C7 MIF haplotype in limited scleroderma but not diffuse scleroderma [75].

Chemokines

Multiple chemokines are involved in the pathogenesis of scleroderma where they mediate chemoattraction of inflammatory cells to the site of dermal fibrosis as well as influence cellular activation, angiogenesis, and fibrosis. Specifically, chemokines including CCL-2 (monocyte chemoattractant
protein-1, MCP-1), CCL-5 (regulated on activation, normal T-cell expressed and secreted, RANTES), CCL-3 (microphage inflammatory protein alpha-1, MIP-\(\alpha\)1), and CXCL-8 (interleukin 8, IL-8) have been shown to be involved in scleroderma [76–79]. For example, lesional and nonlesional scleroderma skin biopsies show increased expression of CCL-2 protein and mRNA, and murine studies have confirmed an important role for CCL-2 in dermal fibrosis [78,80–82].

A small study of 18 patients who had scleroderma investigated CCL-2 polymorphisms and identified an association of the \(-2518\) GG genotype of CCL-2 with scleroderma [83]. This polymorphism is within the promoter region and was further demonstrated to be associated with increased CCL-2 production in vitro [83]. Another recent report in 99 patients who had scleroderma enrolled at Seoul National University Hospital failed to show any significant associations of scleroderma with polymorphisms in CCL-2, CXCL-8, CCL-5, CCL-3, and multiple chemokine receptors [84]. A significant and interesting interaction between polymorphisms of CCL-5 and CXCL-8 and susceptibility to scleroderma was noted using two different statistical approaches, however. These data are intriguing and point to the importance of potential gene–gene interactions in complex diseases such as scleroderma.

**Other candidate genes in autoimmunity/inflammation**

Many additional genetic-association studies of genes involved in autoimmunity and inflammation have been conducted in scleroderma. These reports often have studied small cohorts from different ethnic backgrounds, with conflicting results. These reports are briefly summarized in Table 2 and have been discussed previously in review articles [99,100]. These studies have shown associations of scleroderma with polymorphisms in genes encoding tumor necrosis factor-alpha, cytotoxic T-lymphocyte-associated antigen 4, killer cell immunoglobulin-like receptors, and CD19 receptor, along with others listed in Table 2 [58,88,90,97,98,101,102]. Additional studies are needed to replicate these studies using larger cohorts.

**Vascular function**

**Nitric oxide synthase**

Nitric oxide, a potent vasodilator and critical regulator of vascular tone, is synthesized from \(l\)-arginine by nitric oxide synthase (NOS). Three isoforms of NOS are known: NOS\(_1\) (neural NOS, nNOS), NOS\(_2\) (inducible NOS, iNOS), and NOS\(_3\) (endothelial NOS, eNOS). Multiple studies have shown dysregulation of nitric oxide production (summarized in Ref. [103]). Because of the complexities of nitric oxide functions, it remains unclear if nitric oxide plays a beneficial or pathologic role in scleroderma [103]. Given the importance of nitric oxide in the regulation of vascular tone and vascular endothelial dysfunction in scleroderma, the nitric oxide pathway is an intriguing candidate to consider for genetic-association studies.
Although an initial study in Italians failed to show an association of NOS polymorphisms, a subsequent study of 73 Italian patients who had scleroderma demonstrated that the eNOS 894T allele was associated with disease \[104,105\]. The eNOS 849T allele subsequently was demonstrated to be

Table 2
Candidate genes in inflammation and scleroderma susceptibility

<table>
<thead>
<tr>
<th>Candidate gene</th>
<th>Number of patients</th>
<th>Association</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIF-1</td>
<td>144</td>
<td>A-allele of rs2269475 associated with SSc</td>
<td>[58]</td>
</tr>
<tr>
<td></td>
<td>1015</td>
<td>T-allele of rs2269475 associated with ACA+ SSc</td>
<td>[59]</td>
</tr>
<tr>
<td>CD19</td>
<td>134</td>
<td>−499T allele associated with SSc</td>
<td>[85]</td>
</tr>
<tr>
<td>CD22</td>
<td>126</td>
<td>c.2304 A/A genotype associated with limited SSc</td>
<td>[86]</td>
</tr>
<tr>
<td>CD86</td>
<td>221</td>
<td>−3479G allele associated with SSc</td>
<td>[87]</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>137</td>
<td>+49 AG heterozygotes associated with SSc in African Americans</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>+49A associated with RNP+ SSc only</td>
<td>[89]</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>−1722C, −1661G and −318T (not +49A) associated with SSc</td>
<td>[90]</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>−318T and +49A not associated with SSc</td>
<td>[91]</td>
</tr>
<tr>
<td>CCL-2</td>
<td>18</td>
<td>−2518 GG genotype was associated with SSc</td>
<td>[83]</td>
</tr>
<tr>
<td>CCL-5</td>
<td>99</td>
<td>No association with SSc alone, but a gene–gene interaction with CXCL-8 and SSc was observed</td>
<td>[84]</td>
</tr>
<tr>
<td>CXCL-8</td>
<td>99</td>
<td>No association with SSc alone, but a gene–gene interaction with CXCL-5 and SSc was observed</td>
<td>[84]</td>
</tr>
<tr>
<td>CXCR-2</td>
<td>128</td>
<td>+785 CC and 1208 TT homozygotes associated with SSc</td>
<td>[92]</td>
</tr>
<tr>
<td>IL-1α</td>
<td>86</td>
<td>CTG haplotype association with SSc</td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>IL-1a −889T allele associated with SSc</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td>204</td>
<td>No association with SSc</td>
<td>[51]</td>
</tr>
<tr>
<td>IL-1β</td>
<td>204</td>
<td>No association with SSc</td>
<td>[51]</td>
</tr>
<tr>
<td>IL-2</td>
<td>78</td>
<td>IL-1B31-C and IL-1B-511-T associated with SSc</td>
<td>[52]</td>
</tr>
<tr>
<td>IL-10</td>
<td>140</td>
<td>GCC haplotype underrepresented in SSc</td>
<td>[93]</td>
</tr>
<tr>
<td></td>
<td>161</td>
<td>GCC haplotype associated with SCL70+ SSc</td>
<td>[94]</td>
</tr>
<tr>
<td>IL-13</td>
<td>174</td>
<td>rs18000925 C/T genotype and rs2243204T allele associated with SSc</td>
<td>[95]</td>
</tr>
<tr>
<td>IL-13RA2</td>
<td>97</td>
<td>rs638376 G allele associated with diffuse SSc</td>
<td>[96]</td>
</tr>
<tr>
<td>MIF</td>
<td>486</td>
<td>C7 haplotype associated with limited SSc</td>
<td>[75]</td>
</tr>
<tr>
<td>PTPN22</td>
<td>121</td>
<td>No association with SSc</td>
<td>[69]</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>No association with SSc</td>
<td>[68]</td>
</tr>
<tr>
<td></td>
<td>1120</td>
<td>R620W associated with anti-topo and ACA+ SSc</td>
<td>[70]</td>
</tr>
<tr>
<td>TNF-α</td>
<td>214</td>
<td>−1031C and −863A alleles associated with ACA+ SSc</td>
<td>[97]</td>
</tr>
<tr>
<td></td>
<td>114</td>
<td>−238A allele and +489 AG genotype associated with SSc</td>
<td>[98]</td>
</tr>
<tr>
<td></td>
<td>144</td>
<td>−1031T/T and −237G/G genotypes associated with SSc</td>
<td>[58]</td>
</tr>
</tbody>
</table>

**Abbreviations:** CTLA, cytotoxic T-lymphocyte antigen; IL-13RA2, interleukin 13 receptor alpha 2; TNF-α, tumor necrosing factor alpha.
associated with altered blood-flow profiles in patients who had scleroderma [106]. Additional studies, however, have failed to demonstrate a consistent association of this polymorphism with scleroderma susceptibility [107,108]. Two polymorphisms and a pentanucleotide repeat in \textit{iNOS} were found to be associated with pulmonary arterial hypertension in a cohort of 78 Japanese patients who had scleroderma, but it was not associated with the risk of scleroderma [109]. Consistent with the hypothesis that \textit{iNOS} is decreased in scleroderma, in vitro experiments demonstrated that the pentanucleotide repeat was associated with less \textit{iNOS} transcription [109]. No additional genetic-association studies with \textit{iNOS} polymorphisms have been reported.

\textbf{Angiotensin-converting enzyme}  
Angiotensin-converting enzyme (ACE) catalyzes the conversion of angiotensin I into angiotensin II, which is vasoactive, stimulates aldosterone, and inactivates the vasodilator bradykinin. An insertion/deletion polymorphism of the \textit{ACE} gene is associated with higher tissue and plasma ACE levels (D-allele) [110]. In a study of 73 Italian patients who had scleroderma, the D-allele was associated with scleroderma [105], but this association was not confirmed in a study of 164 American patients who had scleroderma [107].

\textbf{Vascular endothelial growth factor}  
Vascular endothelial growth factor (VEGF) is a major angiogenic factor and prime endothelial cell growth factor. Despite defects in angiogenesis, patients who have scleroderma have high levels of circulating VEGF [111,112]. VEGF polymorphisms have been associated with alterations in VEGF levels and with other diseases, including giant cell arteritis [113,114]. In a genetic-association study of 416 patients who had scleroderma and 249 controls, no differences in allele and genotype frequencies were observed [115].

\textbf{Endothelin-1}  
Endothelin-1 (ET-1) is a potent vasoconstrictive peptide that plays a crucial role in vascular damage. Through interaction with its receptors (ET\textsubscript{A} and ET\textsubscript{B}), ET-1 also regulates tissue remodeling and fibrosing by promoting fibroblast activation and proliferation [116]. Serum and bronchoalveolar lavage fluid from patients who have scleroderma have increased levels of ET-1 that correlate with severity [117,118]. A recent report evaluated the distribution of polymorphisms in ET-1 (\textit{EDN1}), ET\textsubscript{A} receptor (\textit{EDNRA}), and ET\textsubscript{B} receptor (\textit{EDNRB}) [119]. No differences were observed between patients who had scleroderma with controls. Patients who had diffuse scleroderma, however, had an increased frequency of the \textit{EDNRB-1A}, \textit{EDNRB-2A}, and \textit{EDNRB-3G} alleles. Furthermore, the presence of anti-RNA polymerase antibodies in patients who had scleroderma was associated with the \textit{EDNRA} alleles H323H/C and E335E/A [119].
Fibrosis and extracellular matrix production

**Transforming growth factor beta-1**

The importance of TGF-β in scleroderma pathogenesis has been demonstrated and reviewed extensively in the past [120,121]. TGF-β, its receptor, and downstream signaling molecules (eg, Smad3) are expressed at increased levels in affected organs. TGF-β activates dermal fibroblasts leading to increased production of extracellular matrix. Given the importance of TGF-β in scleroderma, it has been hypothesized that polymorphisms in TGF-β may contribute to scleroderma susceptibility. Supporting this hypothesis is a study of 152 patients who had scleroderma that demonstrated an association of the TGFβ +889 C allele with both limited and diffuse scleroderma [122]. These data, however, were not replicated in a subsequent study [123]. More studies with large cohorts as well as studies investigating polymorphisms in the TGF-β receptor, Smad proteins, and non-canonical TGB-β signaling pathways may provide important insight into the role of TGF-β in scleroderma.

**Connective tissue growth factor**

Connective tissue growth factor (CTGF), also known as CCN2, is a heparin-binding 38-kD polypeptide that induces proliferation, extracellular matrix production, and chemotaxis of mesenchymal cells [124–126]. These processes are central to the development of fibrosis of skin and internal organs observed in scleroderma. The production of CTGF is increased in scleroderma skin biopsies and in fibroblasts cultured from patients who have scleroderma [14,127–129]. These observations led to interest in determining if polymorphisms in CTGF were associated with susceptibility to scleroderma [130]. In a cohort of 200 patients who had scleroderma and 188 controls, an association of scleroderma with homozygosity for the G-allele at position –945 of CTGF was observed. Furthermore, this association was replicated in a second cohort of 300 patients who had scleroderma and 312 controls. The odds ratio for the GG genotype using the combination of both cohorts was 2.2 (95% CI, 1.5–3.2). Furthermore, homozygosity for the G-allele was associated with the presence of fibrosing alveolitis and the presence of anti-topoisomerase I antibodies. These data point to an association between the G-945C polymorphism in CTGF and scleroderma susceptibility.

In the same report, the authors extended the initial genetic association using in vitro studies to demonstrate a potential functional relevance of the CTGF G-945C polymorphism in scleroderma [130]. Because this polymorphism occurs in the promoter region of CTGF, the approach focused on the transcription of CTGF. DNA from the CTGF promoter region of five patients who had the CC genotype and five patients who had the GG genotype was cloned into luciferase promoter-reporter constructs. The GG genotype had higher rates of transcription in vitro than the CC genotype. Additional studies demonstrated that Sp1 (an activator of
transcription) bound less efficiently to the CTGF promoter with the GG genotype than with the CC genotype. In contrast, Sp3 (a repressor of transcription) had increased binding to the CTGF promoter with the GG genotype. The shift in the balance of Sp1 binding toward Sp3 binding is consistent with an increase in transcription of the CTGF gene. The combination of the genetic association of the G-945C CTGF promoter polymorphism with scleroderma in two cohorts and the functional data strongly support an important role for CTGF in scleroderma.

Secreted protein, acidic and rich in cysteine

Secreted protein, acidic and rich in cysteine (SPARC) is a matricellular protein in the extracellular matrix that is overexpressed by scleroderma fibroblasts [131]. SPARC regulates cell-to-cell and cell-to-matrix interactions and also is involved in the production of extracellular matrix by scleroderma fibroblasts [132,133]. In a cohort of 20 Choctaw patients who had scleroderma and 178 patients from the GENOSIS cohort, the C allele at +998 position of SPARC was found to be associated with scleroderma across four ethnic populations [131]. Furthermore, fibroblasts isolated from patients homozygous for the +998C allele had more stable expression of SPARC mRNA than fibroblasts from heterozygotes, confirming a functional relevance for the polymorphism that is consistent with the role of SPARC in the pathogenesis of scleroderma [131]. Despite the potential functional importance of this polymorphism in SPARC, a subsequent study was unable to confirm a genetic association with susceptibility to scleroderma [134].

Genome-wide studies

Genome-wide association studies have been a powerful approach to the identification of genetic susceptibility loci in complex genetic disorders such as rheumatoid arthritis, inflammatory bowel disease, and coronary artery disease [135–138]. Genome-wide association studies allow the investigator to identify new mechanisms and pathways without predefined assumptions or biases. Such studies rely on large cohorts of patients who have well-defined disease phenotypes. The only genome-wide association study in scleroderma was performed in the Choctaw Indians, who have the highest known prevalence of scleroderma [139]. This study used 400 microsatellite markers to identify 17 chromosome regions associated with susceptibility to scleroderma, including the HLA region and regions near the SPARC, FBN1, and topoisomerase I (TOPOI) genes. These data must be replicated in other populations, and better resolution of the genes within these chromosomal regions is needed. With recent technological advances, high-density genotyping experiments using SNP arrays to study DNA variations are increasingly reported. Current platforms are able to interrogate more than 1 million SNPs on a single microarray. These high-density, large,
genome-wide association studies in patients who have scleroderma are currently underway and will be important in defining the future direction of disease association as well as mechanistic research in scleroderma.

**Microarray studies**

Several technologies that have emerged since the sequencing of the human genome are particularly useful in dissecting the molecular mechanisms of complex biologic processes. Gene-expression profiling using whole-genome microarrays is one such powerful technique that allows simultaneous analysis of thousands of mRNA transcripts. The investigator can learn about multiple pathways involved in diseases and discover completely novel pathways that previously would not have been hypothesized to be involved. This approach also enables the investigation of complex biologic processes in a relatively unbiased manner. These studies, however, must use rigid statistical approaches and must use alternative methodologies to confirm the results. With these requirements in mind, several studies have used microarray expression profiling in scleroderma, and the data that are emerging certainly will guide the direction of future research.

Patterns of gene expression were determined in biopsies of lesional and nonlesional skin from four patients who had scleroderma and four healthy controls [140]. Initial analyses demonstrated that expression of 2776 genes varied by more than twofold, and unsupervised hierarchical clustering grouped all but one of the scleroderma samples into a single group. This group included both lesional and nonlesional skin biopsies from patients who had scleroderma. Clusters of genes that differed among patients who had scleroderma and healthy controls were groups of genes expressed in endothelial cells, smooth muscle cells, T lymphocytes, and B lymphocytes, as well as genes involved in extracellular matrix. No differences in gene-expression profiles were seen in the small number of cultured normal or scleroderma fibroblasts studied [140,141].

A more recent study compared the gene-expression profiles of skin biopsies and corresponding cultured fibroblasts from nine patients who had scleroderma and nine controls [142]. Unsupervised clustering of all skin biopsy samples demonstrated that the scleroderma phenotype was the dominant influence on the expression profile. Expression profiles of scleroderma and control biopsies differed in 1839 genes ($P < .01$). Alterations in TGF-$\beta$ and Wnt pathways, extracellular matrix proteins, and the CCN family were notable in scleroderma skin biopsies. A key finding of these data is that scleroderma in the skin has a distinct and complex gene-expression profile that is reflected only partially in matched explanted fibroblasts cultured under standard conditions. Despite the large numbers of genes differentially expressed in scleroderma skin biopsies, the expression profile from the cultured fibroblasts differed only in the expression of 223 genes ($P < .01$). These data emphasize the importance of studying skin
biopsies and also suggest a significant unmet need to develop better in vitro systems for the study of the complex molecular mechanisms of scleroderma. One such model has been reported recently in which normal dermal fibroblasts are transfected with adenovirus vector expressing TGF-β receptor-1 \([143,144]\). The overexpression of TGF-β receptor-1 in dermal fibroblasts causes alterations in the gene-expression profile that are similar to those of scleroderma fibroblasts \([143,144]\).

Dysregulation of the immune system is central to the pathogenesis of scleroderma. Prior investigations of circulating immune cells have focused largely on individual populations of cells, such as T lymphocytes. These studies are critical to advancing the understanding but are limited in that they are unable to capture changes that may be observed when multiple populations of cells are allowed to interact with each other. To obtain a global view of the differences in gene-expression profile in early scleroderma, total RNA harvested from peripheral blood cells (PBCs) from 18 patients who had untreated, diffuse scleroderma with recent disease onset and 18 matched controls was analyzed using transcriptional profiling \([145]\). Global analyses demonstrated an increase in the expression of 244 genes and decreased expression of 138 genes in scleroderma PBCs compared with controls. These studies demonstrated differential expression of 18 interferon-α-inducible genes, including six genes that also are altered in lupus peripheral blood mononuclear cells. Recently, this interferon signature has been confirmed in another microarray study of scleroderma peripheral blood mononuclear cells \([146]\). Furthermore, this report demonstrated that activation of interferon-regulated genes was dependent on the Toll-like receptors, TLR-7 and TLR-9 \([146]\). In addition to interferon-regulated genes, microarray analyses of PBCs from patients who had scleroderma demonstrated increased expression of AIF-1 and cellular adhesion molecules including selectin and integrin family members \([145]\). In total, 13 biologic pathways were differentially regulated in scleroderma PBCs, including GATA-3, a key transcription factor involved in T-helper cell type 2 differentiation \([145]\). These data are available for review online at http://www.uth.tmc.edu/scleroderma. Taken together these reports demonstrate that transcriptional profiling can discriminate reliably between scleroderma and control PBCs and will be useful in the identification of hitherto unrecognized disease subsets and in the discovery of molecular pathways involved in scleroderma.

**Summary and future directions**

The evidence for a strong genetic contribution to scleroderma pathogenesis continues to mount. Epidemiologic studies suggest that a family history is the strongest risk factor, but ethnicity also contributes. The candidate-gene association studies presented in this article identify several individual genes that may be involved in scleroderma. Biologic pathways do not act...
in isolation, however. More often than not, any given pathway will have significant crosstalk with a multitude of other pathways through common intracellular signal transduction molecules and transcription factors. Indeed the whole-genome microarray expression studies of skin biopsies and peripheral blood cells from patients who have scleroderma, as described in previous sections, supports the involvement of multiple pathways in the development of scleroderma. The end result is a complex regulatory network of genes and pathways that, because of limitations in knowledge, initially may appear counterintuitive. Future studies need to foster the understanding of the genes and proteins involved in these pathways and how they interact with one another to lead ultimately to the development of scleroderma.

Although the gene-association studies reviewed provide insight into the development of scleroderma, they have significant limitations that can be addressed in future studies. Many candidate-gene association studies used small cohorts of patients who had scleroderma and controls. Population stratification and differences in linkage disequilibrium make interpretation of the data from these studies difficult. To help address these limitations, it is imperative that future studies be performed in large cohorts of patients to help determine the genetic effects that are involved in scleroderma. Because of the low prevalence of scleroderma, extensive multi-institutional collaborations will be necessary to develop these large cohorts. These collaborations also will enable the formation of cohorts for replication studies, a critical step in confirming initial observations.

The clinical heterogeneity of scleroderma is a substantial obstacle to the design and interpretation of studies. Although large sample sizes are essential, future cohorts also must be able to characterize the patients carefully according to demographics, clinical phenotype, autoantibody profiles, and clinical course. Associations between distinct scleroderma autoantibody profiles and clinical manifestations may allow better definition of scleroderma disease subsets. Indeed, as discussed previously, stratification of patients based on autoantibody profiles demonstrates stronger associations with HLA polymorphisms and with other candidate genes (Table 3).

Another significant limitation of the candidate-gene approach that has been used to date is the inability to investigate large numbers of genes concomitantly. The completion of the human genome, advances in the understanding of the genome, and progress in the development of high-throughput molecular and genetic technologies have facilitated the use of genome-wide association studies to study genetic determinants of complex diseases. Although these studies require large cohorts and sophisticated statistical analyses, they can be performed without bias in terms of the genes that are considered for analysis. In addition, genome-wide approaches will make possible the identification of completely novel genes involved in scleroderma susceptibility that would not have been considered using the traditional candidate-gene approach. This approach now is being used
<table>
<thead>
<tr>
<th>Autoantibody</th>
<th>Clinical characteristics</th>
<th>HLA associations</th>
<th>Candidate-gene associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-topoisomerase I</td>
<td>Diffuse skin involvement</td>
<td>Whites: DRB1<em>1104, DPB1</em>1301</td>
<td>PTPN22 R620W</td>
</tr>
<tr>
<td></td>
<td>Pulmonary fibrosis</td>
<td>African Americans: DQBI<em>0301, DPB1</em>1301</td>
<td>CTGF G945C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Japanese: DRB1<em>1302, DQBI</em>0601, DPB1*0901</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Choctaw: DRB1<em>1602, DQBI</em>0301, DPB1*1301</td>
<td></td>
</tr>
<tr>
<td>Anti-centromere</td>
<td>LSSc</td>
<td>DQBI*0501 and other DQBI alleles having polar amino acids in position 26</td>
<td>AIF-1 +899T</td>
</tr>
<tr>
<td>Anti-RNA polymerase</td>
<td>Pulmonary hypertension; digital ulcers</td>
<td></td>
<td>PTPN22 R520W</td>
</tr>
<tr>
<td>Anti-PM-Scl</td>
<td>Renal crisis</td>
<td>DQBI*0201</td>
<td>EDNRA H323/H and E335E/A</td>
</tr>
<tr>
<td>Anti-U3RNP (fibrillarin)</td>
<td>Rapid progressive skin fibrosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overlap myositis</td>
<td>DRB1*0301</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Young age of onset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African Americans and males</td>
<td>DRB1<em>1302, DQBI</em>0604/0605</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviation:** LSSc, limited scleroderma.
extensively and already has provided significant insight into diseases such as rheumatoid arthritis and systemic lupus erythematosus [135,136].

Following the identification of individual genetic determinants of scleroderma susceptibility, it will be necessary to increase the understanding of how these genetic polymorphisms relate to the development of scleroderma. Central to this understanding will be defining complex gene–gene and gene–environment interactions. In addition, biologic confirmation of these genetic alterations into functional studies is essential to determine whether these associations are, in fact, causal. Such studies will depend on the availability of model systems for scleroderma that are amenable to experimental manipulation and accurately reflect the disease phenotype. Last, efforts also should focus on translating these findings into the care of the patient who has scleroderma. Ultimately, these genetic factors may have great importance in the diagnosis, prognosis, and perhaps treatment of patients who have scleroderma.

References


