Review

Genetic background of systemic sclerosis: autoimmune genes take centre stage

Yannick Allanore¹,², Philippe Dieude³ and Catherine Boileau²,⁴

Abstract

SSc is a complex multiorgan disease. The key steps in its pathogenesis include early endothelial damage, dysregulation of the immune system with abnormal autoantibody production and fibroblast activation resulting in hyperproduction of extracellular matrix. The disease is caused by an interaction between susceptibility genes and environmental triggers since epidemiological data, including family and twin studies, reveal a genetic component in the pathogenesis of SSc. The candidate gene approach has mainly been employed to identify SSc susceptibility genes. We will focus on data obtained through large samples of well-phenotyped patients and replicated in independent cohorts. These case–control association studies have enabled the identification of several genes that are shared with other connective tissue disorders, and for some of these, putative autoimmune susceptibility genes have been identified. Indeed, we will mainly focus on IRF5 (rs2004640), STAT4 (rs7574865), PTPN22 (rs2476601) and BANK1 (rs3733197 and rs10516487) data. Some of these genes/loci are common to several autoimmune diseases, indicating a shared genetic background also contributing to SSc. Among connective tissue disorders, similarities for genetic markers with SLE are noteworthy. Most likely, these immune-modifying genes could interact and influence both disease phenotype and severity. Less evidence is available yet with regard to genetic markers relating to the vascular and fibrotic aspects of the disease.

Key words: SSc, Genes, Polymorphisms, Autoimmunity.

Introduction

SSc is a complex multiorgan disease affecting the immune system, the microvascular network and the connective tissue. Only few aspects of the disease pathogenesis are known, starting from the early inflammatory phase to the fibrosis of cutaneous and internal organs. Progressive organ failure renders SSc a severe and lethal disease: it may display an acute evolution that may become chronic leading to a major disability and loss of quality of life and even death. Many points have been clarified by recent epidemiological studies [1–5]. It has been confirmed that SSc is the most severe form of CTD: the standardized mortality ratio was estimated at 1.5–7.2 in an international study [2] and at 2.7 in Canada [3]. Several organ involvements remain untreatable and no specific drug counteracting the fibrotic process has been found. This has been highlighted by the poor results recently reported in SSc-related pulmonary arterial hypertension [6] and interstitial lung disease [7].

Prevalence of the disease varies worldwide and population-based studies yield higher prevalence than hospital records-based studies. In North America, the prevalence of SSc has been reported as 74.4 cases per 100 000 women [95% CI (confidence interval) 69.3, 79.7] and 13.3 cases per 100 000 men (95% CrI 11.1, 16.1) in a Canadian study [8], whereas in the USA figures of 26 per 100 000 [9] vs 75 per 100 000 [10] were reported by medical records- vs population-based studies, respectively. In Europe, the prevalence was reported as 16 per 100 000 in Denmark [11] and 15.8 per 100 000 adults (95% CI 129, 187) in France [12].

The scope of this review is to present and discuss the more relevant data on genetic susceptibility factors obtained through large samples and replicated in independent cohorts. We will focus on genes regulating the
immune system (i.e. innate immunity, cell signalling and T-cell biology) for which the data are more consistent. We will discuss shared genetics and mechanisms with other connective tissue disorders and autoimmune diseases, and also the interaction of environment with host susceptibility genes.

**Genetic background**

SSc is not inherited in a Mendelian fashion, and although SSc pathogenesis is unclear, it is believed that both genetic and environmental factors contribute to disease susceptibility and clinical expression or progression [13]. Complex genetic diseases are influenced by the interplay of multiple genes and/or the environment; susceptibility genes act in concert to increase an individual’s risk of disease. Thus, in contrast to the situation in monogenic traits, most susceptibility genes exert only a minor individual effect on the disease itself.

Two complementary analytical methods (linkage with positional cloning and association studies) have been employed to detect the specific genetic regions involved in the disease process. However, in complex diseases, due to the large number of loci that may be involved, and the genetic heterogeneity underlying the phenotypic heterogeneity, it has often been difficult to replicate linkage evidence. The availability of genome-wide markers, combined with multipoint analyses using closely linked markers, has increased the power of studies. The genome-wide association (GWA) approach has yielded a wealth of new genetic susceptibility genes/loci in complex genetic disorders. The completion of the Human Genome Project [14] and high-throughput genotyping technology have enabled researchers to carry out case-control association studies on thousands of samples with several thousands of markers throughout the human genome. This case-control approach is designed to find the loci that fit with the common disease-common variant hypothesis of human disease. It has provided some important results in immune-mediated diseases, such as RA, SLE [15] and Crohn’s disease [16]. However, despite extensive efforts, only a few susceptibility genes have been identified to date.

The risk of SSc is increased among first-degree relatives of patients, compared with the general population. In a study of 703 families in the USA, including 11 multiplex SSc families, the familial relative risk in first-degree relatives was ~13, with a 1.6% recurrence rate, compared with 0.026% in the general population [17]. The sibling risk ratio was ~15 (range 10–27 across cohorts). Although a family history of SSc is the highest risk factor identified to date, the likelihood of developing the disease is <1% among offsprings of patients and others. Furthermore, familial associations might relate to some environmental exposure. The only twin study reported to date included 42 twin pairs (24 monozygotic) [18]. The data show a similar concordance rate in both the monozygotic (4.2%) and dizygotic twins (5.6%; NS), and an overall cross-sectional concordance rate of 4.7% (Table 1). However, concordance for the presence of ANAs was significantly higher in the monozygotic twins (90%) than in the dizygotic twins (40%). It is noteworthy that for one concordant pair, the onset of disease in the first twin was ~15 years after the onset of disease in her sister. In addition, the mean age at evaluation of the monozygotic twins was 48 years and that of the dizygotic twins was 48.5 years (range for both 28–69 years), which is commonly recognized as the age at risk for occurrence of SSc and thus one may argue that a longer follow-up is required before drawing a definitive conclusion for clinical concordance. A further study showed a high level of concordance among monozygotic twins regarding the gene expression profile of cultured dermal fibroblasts [19]. These results might support the role of the genetic background, at least for some component of the disease, but also clearly confirm that genetic susceptibility alone is not sufficient for the development of SSc.

**Familial and individual aggregation of SSc with other autoimmune diseases**

One recent study attempted to ascertain whether a higher risk of developing rheumatic diseases in first-generation immigrant parents in Sweden is associated with a higher risk of developing rheumatic diseases in the next generation [20]. Polish-born immigrants and second-generation Yugoslavs and Russians showed a significantly increased risk of SSc, consistent with the genetic component of the disease. However, this cannot rule out that environmental exposure is relevant for these individuals who may have been exposed to similar environmental factors.

History of the co-occurrence of other autoimmune diseases (AIDs) in SSc patients as well as a family history of autoimmunity has been the subject of several other recent studies. Polyautoimmunity (i.e. AID co-occurring within patients) and familial autoimmunity (i.e. diverse AID co-occurring within families) have been investigated [21]. Among 719 SSc patients, 273 (38%) had at least one other AID according to a cross-sectional study. The most frequent autoimmune diseases were thyroid disease (38%), RA (21%), SS (18%) and primary biliary cirrhosis (4%). There were 260 (36%) SSc patients with their first-degree relatives affected by at least one AID, of which the most frequent were RA (18%) and thyroid disease (9%) [21]. We have confirmed these findings in European Caucasian patients. A study on these patients reported

**Table 1 Genetic background of SSc: evidence from epidemiological data and prevalence analyses**

<table>
<thead>
<tr>
<th>Population</th>
<th>Prevalence, %</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population:</td>
<td>0.026</td>
<td>[16]</td>
</tr>
<tr>
<td>unrelated individuals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family approach:</td>
<td>1.6</td>
<td>[16]</td>
</tr>
<tr>
<td>first-degree relatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monozygotic twins</td>
<td>4.7</td>
<td>[17]</td>
</tr>
</tbody>
</table>
that about a quarter of a large series including 1132 SSc patients developed at least one associated AID [22]. In addition, we observed that the co-existence of at least one AID was associated with the limited cutaneous subtype and a lower prevalence of digital ulcers. Therefore, this latter study identified a subset of patients with milder disease, potentially associated with a specific B-cell-mediated autoimmunity.

Another study has been undertaken to estimate associations of RA with any of 33 AIDs and related conditions among parents and offspring, singleton siblings, twins and spouses [23]. Familial standardized incidence ratio (SIR) for RA in relation to other AIDs and related conditions according to the presence of SSc in the proband were 1.65 (1.13, 1.33) for parents only and 1.99 (0.87, 4.32) for siblings only. For discordant diseases, offspring had an increased risk of RA when parents were hospitalized for SSc (SIR 1.65). These associations support the concept of a shared genetic background.

**Susceptibility genes in SSc: autoimmunity**

The immune system plays a crucial part in the host defense against harmful antigens and in the balance between tolerance and immunity to other antigens. A very large number of associations between the HLA system and autoimmune disorders has long been established [24]. In addition, through the candidate gene approach and more recently by genome-wide scans, a large number of non-HLA candidate genes have been identified including those that have consistently been found to be associated with multiple autoimmune disorders [24]. These results inspired several investigations in SSc and our review aims to discuss the more relevant results obtained, focusing on those that could be replicated in independent populations.

In AIDs, the only unambiguous susceptibility loci identified is the HLA complex, located on the short arm of chromosome 6. Indeed, the MHC region has been the most prominently associated region of the human genome, with different allelic variants associated with each disease. Previous HLA genetic association studies of SSc patients have suggested that these genes exert their influence primarily on autoantibody expression [25, 26]. ACA-positive and anti-topo I antibody-positive SSc subsets have been associated with different HLA class II alleles and haplotypes. HLA-DRB1*11–DQB1*0301 haplotypes have been associated with anti-topo I positivity in SSc, whereas HLA-DRB1*01–DQB1*0501 haplotypes are more common in ACA-positive SSc patients [27]. A recent case–control association study was conducted to determine the HLA class II (DRB1, DQB1, DQA1 and DPB1) alleles, haplotypes and shared epitopes in 1300 SSc cases (961 whites) and 1000 controls in the US population [27]. The strongest positive class II associations with SSc in whites and Hispanics were the DRB1*1104, DQA1*0501, DQB1*0301 haplotypes and DQB1 alleles encoding a non-leucine residue at position 26 (DQB126 epi), whereas the DRB1*0701, DQA1*0201, DQB1*0202 haplotypes and DRB1*1501 haplotype were negatively correlated and possibly protective in dominant and recessive models, respectively. However, several other alleles/haplotypes which appear to be unique to several autoantibody-positive groups were observed, highlighting the influence of ethnic diversity and SSc clinical heterogeneity. This will need to be replicated and clinical translation investigated.

The highly complex linkage disequilibrium (LD) structure of this locus has limited the dissection of the risk variants responsible for the considerable involvement of the HLA locus in AID susceptibility. One particular point for a subset of MHC haplotypes is that there is a higher amount of LD observed between segments of strong LD [28, 29]. Therefore, such tight segment-to-segment LD limits MHC research: if one identifies a disease association with a variant in the region, it may not be possible to determine whether the variant is causal or whether its association simply reflects LD with the true causal variation. A further complication is the great variability exhibited by some of the HLA genes [29]. Studies performed in much larger cohorts that evaluate the entire HLA (classes I, II and III) locus or a dense map of variation rather than specific regions, and that are inclusive of non-Europeans, could be helpful in understanding the contribution of HLA in SSc pathogenesis.

Regarding non-HLA genes, we present below the more consistent results and in particular those that were obtained through large samples and were replicated. According to a recent review on shared autoimmunity [24], we have classified SSc susceptibility genes according to their involvement in specific pathways: (i) innate immunity, (ii) immune-cell signalling regulation and (iii) T-cell differentiation.

**Genes involved in innate immunity**

A growing body of evidence has provided a new paradigm for understanding autoimmunity in CTDs and particularly SLE [30, 31]. Type I IFN is a central mediator of innate immunity and IFN production is normally a feature of the immune response to microbial infections and has multiple effects on the immune system. IFN stimulates monocyte maturation into dendritic cells, plasma cell maturation and immunoglobulin class switching, cytotoxic T- and natural killer cell activity, and chemokine secretion. Microarray studies have underlined a pivotal role of the type I IFN in SLE and primary SS and these data have been strengthened by the identification of the IFN regulatory factor 5 (IRF5) gene, and particularly IRF5 rs2004640, as a susceptibility genetic factor in these diseases [30, 31]. The IRF5 rs2004640 T allele, which creates a donor splice site in intron 1 of the gene, results in the transcription of the alternative exon 1B of no clear biological effect. Gene expression profiling is another complementary genetic analytical method as changes in expression are at least in part related to genetic regions involved in the disease. Several studies have demonstrated its importance in SSc, its potential in prognostication, as well as in the
elucidation of the underlying pathogenesis [32]. Using peripheral blood cells, a type 1 IFN signature has been reported in SSc [33] and raised the hypothesis of a role of IFN genes in this disorder.

An association of IRF5 was shown from a case–control study of our group (1641 subjects of French European Caucasians split into discovery and replication cohorts) [34]. In both sets, the TT genotype was significantly more common in patients with SSc than in control subjects, with an OR of 1.58 (95% CI 1.18, 2.11) in the combined analyses. Among the whole SSc population, we found a significant association between homozygosity for the T allele and the presence of fibrosing alveolitis (OR 2.07; 95% CI 1.38, 3.11). In addition, in a multivariate analysis model including the diffuse cutaneous subtype of SSc and positivity for anti-topo I antibodies, the IRF5 rs2004640 TT genotype remained associated with fibrosing alveolitis ($P$ = 0.029; OR 1.92; 95% CI 1.07, 3.44). Preliminary data from the US cohort, not yet published, have confirmed these findings [35]. Logistic regression analysis showed in this population that the TT genotype of single nucleotide polymorphism (SNP) rs2004640 was an independent risk factor for SSc (OR 1.56; 95% CI 1.3, 2.0) and also for anti-topo I antibody positive and SSc-related fibrosing alveolitis [35]. A replication of significance has been reported in an independent population from Japan [36]. A case–control association study (281 SSc patients and 477 healthy controls) was performed for rs2004640 as well as for rs10954213 and rs2280714. All three SNPs were significantly associated with SSc, with the rs2280714 A allele having the strongest association (OR 1.42; 95% CI 1.15, 1.75). Association was preferentially observed in subsets of patients with dcSSc and anti-topo I antibody positivity. Conditional analysis revealed that rs2280714 could account for most of the association of these SNPs. The genotype of rs2280714 was strongly associated with IRF5 mRNA expression. This replication is noteworthy as it is obtained in a different ethnic group. Altogether, these data provide new insight into the pathogenesis of SSc, including clues to the mechanisms leading to specific disease subtypes. Nevertheless, considering discrepancies between SLE and SSc regarding a fibrotic component, no unifying hypotheses can be put forward. Further research into the role of IRF5 and IFN in promoting specific fibrotic phenotype expression in SSc is warranted.

**Genes involved in T-cell differentiation**

Recent GWA studies for SLE have also identified new associations with genes in T- and B-cell pathways [14, 37]. Indeed, various association studies have convincingly identified the STAT4 gene as a susceptibility factor for different connective tissue disorders and inflammatory bowel diseases [14, 38, 39]. STAT4 encodes the signal transducer and activator of transcription 4, which transduces IL-12, IL-23 and type 1 IFN cytokine signals in T cells and monocytes. The identified STAT4 susceptibility gene variant is rs7574865, located in the third intron of the gene. Although it is a non-functional variant, it tags an LD block which is correlated with high expression of STAT4 mRNA [39]. Two studies have concomitantly identified the association between STAT4 rs7574865 and SSc. One study included a total of 1317 SSc patients and 3113 healthy controls, from an initial case–control set of Spanish Caucasian ancestry and five independent cohorts of European ancestry [40]. The rs7574865 T allele variant was significantly associated with susceptibility to lcSSc in the Spanish population (OR 1.61; 95% CI 1.29, 1.99), but not with dcSSc. The combined analysis showed a strong risk effect of the T allele for lcSSc susceptibility (pooled OR 1.54; 95% CI 1.36, 1.74). The other study included 1855 individuals, all of French Caucasian origin [41]. STAT4 rs7574865 was found to be associated with SSc (OR 1.29; 95% CI 1.11, 1.51). However, this association was not restricted to a particular phenotype and an association with fibrosing alveolitis was observed. A case–control association study has been performed in the Japanese population and included 282 patients with SSc and 590 controls. In spite of the difference in the risk allele frequency between the populations, the data confirmed an association between rs7574865 and SSc, preferentially with lcSSc subset [42]. Therefore, although these three studies established STAT4 as a genetic risk factor, genotype–phenotype correlations remain unclear and will need further research. In our study, we also investigated the joint effect on SSc susceptibility of the risk allele IRF5 rs2004640 T and STAT4 rs7574865 T [41]. No evidence for dominance or interactions was observed between the two risk alleles. We observed an additive effect on SSc susceptibility. The ORs for SSc were 1.72 (95% CI 1.23, 2.14) for carriers of one risk allele; 2.17 (95% CI 1.55, 3.04) for carriers of two risk alleles; and 2.72 (95% CI 1.86, 3.99) for carriers of three or four risk alleles. We also observed a strong association between the three risk alleles and SSc interstitial lung disease (OR 1.79; 95% CI 1.235, 2.582), which remained significant in multivariate analyses including other known risk factors for lung involvement. This point is of major interest as an additive effect has been reported in an SLE Caucasian population, for contribution to a general loss of tolerance [43]. Hence, STAT4 and IRF5 could both contribute to a disease-specific phenotype.

**Genes involved in immune-cell signalling**

PTPN22 is a selective phosphatase that modulates signal transduction in T cells, and represents a case in which a causal variant that contributes to disease susceptibility has been identified. The known p.R620W (1856C to T; rs2476601) risk allele is a gain of function variant with increased catalytic activity, and is thought to be a more potent suppressor of TCR signalling [24, 43]. The PTPN22 gene has been convincingly associated with Crohn’s disease, RA, SLE, type 1 diabetes and Graves’ disease [44, 45]. This clustering of genetic risk factors for many AIDs shows that these diseases may share, at least partly, similar underlying causal mechanisms. In SSc, the first evidence of association of PTPN22 was reported by a case–control study of >1000 SSc patients in
the USA [46]. The PTPN22 CT/TT genotype showed significant association with anti-topo I antibody-positive SSc in white patients (OR 2.21; 95% CI 1.3, 3.7) and with ACA-positive white patients with SSc (OR 1.70; 95% CI 1.1, 2.7). Frequency of the T allele also showed significant association with anti-topo I antibody-positive SSc in white patients (OR 2.03; 95% CI 1.3, 3.2). Replication came from performing a case–control study (659 SSc patients and 504 healthy controls) enlarged to additional SNPs (n = 7) and by the performance of a meta-analysis [47]. In a first step, we assessed our SSc patients for concomitant AID and excluded those with known PTPN22 risk allele-associated diseases. No association was detected for any of the SNPs tested. However, PTPN22 haplotype analysis identified a strong association between SSc and the presence of a risk haplotype carrying the 1858T allele (P = 1.52 × 10−10) and a protective haplotype carrying the 1858C allele (P = 2.20 × 10−11). The meta-analysis provided further evidence that the PTPN22 1858T allele is involved in genetic susceptibility to SSc in Caucasian (OR 1.08; 95% CI 1.02, 1.15) and mixed (OR 1.09; 95% CI 1.04, 1.16) populations; particularly in the anti-topo I-positive subset (OR 1.48; 95% CI 1.09, 2.0 for Caucasians and OR 1.09; 95% CI 1.04, 1.16 for mixed). Altogether, these results driven by large studies indicate that PTPN22, a shared genetic factor of multiple autoimmune diseases, also contributes to the genetic background of Ssc. Regarding the overall data obtained with PTPN22, a role for genetic factors in autoantibody sero-positivity has emerged. Hence, autoantibody positivity might represent sub-phenotypes within autoimmune diseases, and differential genetic associations in autoimmune-positive and -negative groups could help us to understand differences in pathogenesis and disease development.

Scaffold protein with ankyrin repeat gene, BANK1, is a B-cell-specific scaffold protein and LYN tyrosine kinase substrate that promotes tyrosine phosphorylation of inositol 1,4,5-trihphosphate receptors. BANK1 overexpression leads to the enhancement of BCR-induced calcium mobilization by connecting protein tyrosine kinases to IP3R, facilitating phosphorylation and activation of IP3R by LYN and subsequent release of Ca2+ from endoplasmic reticulum stores. Nevertheless, to date, the exact role of BANK1 remains unclear. Functional variants of BANK1 were recently found to be associated with SLE [48]. Therefore, we tested this newly identified risk factor for SLE for association with Ssc, and for its putative gene–gene interaction with the pre-identified genetic risk factors IRF5 and STAT4 [49]. BANK1 non-synonymous functional substitution rs10516487 and rs3733197 SNPs were genotyped in 2407 individuals comprising a French set (874 SSc patients and 955 controls, previously genotyped for both IRF5 rs2004640 and STAT4 rs7574865) and a German set (421 SSc patients and 182 controls). The BANK1 variants were found to be associated with dcSSc in both samples, providing an OR of 0.77 (95% CI 0.64, 0.93) for the rs10516487 T rare allele in the combined populations compared with controls and an OR of 0.73 (95% CI 0.61, 0.87) for the rs3733197 A rare allele. BANK1 haplotype analysis found the A-T haplotype to be protective in dcSSc (OR 0.70; 95% CI 0.57, 0.86); significant differences were also observed when the lcSSc subset was compared with dcSSc, both for rare alleles and haplotypes. Moreover, the BANK1, IRF5 and STAT4 risk alleles displayed a 1.43-fold increased risk of dcSSc. These results suggest a role of B cells in the dcSSc subset of patients, in accordance with the findings for the B-cell stimulator, BAFF, which showed higher serum levels in dcSSc patients that were correlated with the skin score [50]. It is noteworthy that we could identify no link with autoantibody status. Similarly, a recent study in RA that reported a modest effect at the BANK1 locus could identify no link with specific autoantibodies [51].

Although autoantibodies are supposed to be of greater importance in SLE, a relationship with BANK1 variants has not yet been queried in this setting. We are reaching a new area in complex genetic diseases where identification of the genes involved in disease susceptibility or severity is becoming a reality. Therefore, it is of interest to understand the relationship between the various genetic effects. We examined the possibility of a genetic interaction between a B-cell genetic marker, BANK1, and type I IFN genes, IRF5 and STAT4. Our results suggested that the IRF5 and STAT4 SNPs act additively with BANK1 to increase the risk for dcSSc and SSc [49]. Therefore, at least two major pathways seem to contribute to dcSSc pathogenesis, implicating both innate and adaptive immunity (associated autoimmune genes are summarized in Table 2).

**Susceptibility genes in SSc: outside autoimmunity**

Several studies have been performed during the same period to identify variants from vascular or fibrotic candidate genes that may be associated with SSc. Despite their major relevance from a pathogenetic point of view and the use of large studies in some cases, no gene could be consistently identified and replicated [13, 52]. Probably the most important data came from the CTGF gene. A functional SNP, at –945 bp from the start codon in the promoter region (rs6918698), was found to be associated in a UK population including 1000 subjects [53]. However, other studies of larger sample size (2964 individuals of European ancestry and 2315 from the USA) failed to replicate these findings [54, 55]. It is noteworthy that overall these studies reported similar allele or genotype frequencies in the SSc group but discrepancies regarding the control group that supported the association we found in the initial report. However, a study from Japan including 659 individuals has recently suggested an association with the same SNP [56]. Therefore, the definitive involvement of CTGF variants in the genetic background of SSc remains to be established and new studies and/or meta-analyses remain to clarify this issue.
Future directions

These genetic studies indicate that certain genes/loci seem to confer predisposition to multiple immune-related disorders, thereby supporting the hypothesis that a shared group of genes contributes to the spectrum of immune diseases. They confirm the clinical and epidemiological observations regarding the co-occurrence of immune-related diseases. They also provide clues indicating why these immune diseases form clusters. Both the innate and adaptative immunities appear to play a role in these diseases. Innate immunity is particularly interesting from a clinical perspective, as it provides links to environmental triggers of the disease. It is noteworthy that despite their pleiotropic effects, these genes seem to be associated with some severe sub-phenotypes of the immune disorders. This situation may lead in the future to reconsideration of the classification of some diseases according to a better understanding of their pathogenesis.

An important issue is the research agenda from the current situation. It is important to correlate genotypic and phenotypic data that may provide insight into disease pathogenesis and, more importantly, new therapeutic approaches. It will be important to determine whether the genetic variants identified to date fall into specific mechanistic pathways or involve a multitude of different pathways. There is a great need for post-genomic studies paired with comprehensive genotypic data. Indeed, the identification of the different molecular pathways involved may provide innovative targets for therapeutic intervention. New therapies are particularly expected in SSc which still has no specific efficient therapy. Longitudinal studies will be instrumental in assessing the contributions of both genetic and environmental factors. They may also allow prediction of disease progression. The genetic predisposition to impairment of a given pathway could help define clinical subgroups of the disease and prioritize patient groups towards a specific therapy. To that end, unified efforts from clinicians and scientists of complementary expertise will be mandatory.

In SSc, no GWAS study has been published yet and one may expect important information to be driven through this approach. It is likely that many different susceptibility alleles contribute to SSc, each of which has only a modest effect. Therefore, one may speculate that many genes remain to be identified. Genes or more probably pathways that could act at the interplay between the respective disturbances will be particularly scrutinized. An advantage of these approaches, as compared with more typical association studies that test for a connection between disease and candidate-gene markers, is that they screen most of the genes in the human genome—thus allowing the investigation of new mechanisms, sometimes unexpected, of disease susceptibility. Nevertheless, the heterogeneity of SSc may limit somehow the establishment of true associations and this will need careful multistep statistical analyses. One may hope that GWAS will identify not only a number of genome-wide significant associations but also additional sets of ‘suggestive’ findings that will have to be followed up and replicated in other independent populations. Thereafter, meta-analyses will be required to identify associations. In addition, examining rare variants (rarer variants of larger effects) and copy number variants may provide further insight into genetic mechanisms of complex diseases to which SSc belongs. Finally, and in contrast to genetic alterations, an epigenetic change is defined as a heritable change in gene expression that does not involve a change in the DNA sequence. Epigenetic mechanisms play an essential role in eukaryotic gene regulation by modifying the chromatin structure, which in turn modulates gene expression. Convincing evidence indicates that epigenetic mechanisms, and in particular impaired T-cell DNA methylation, provide additional factors contributing to connective tissue disorders [57] and this will also have

**Table 2** Autoimmunity susceptibility genes associated with SSc

<table>
<thead>
<tr>
<th>Marker</th>
<th>Gene/locus</th>
<th>Chromosomal region</th>
<th>Relevance to pathogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs 2004640 (exon 1B, alternative splicing)</td>
<td>IRF5</td>
<td>7q32</td>
<td>Member of the IRF family; transcription factors with diverse roles, including modulation of cell growth, differentiation, apoptosis and immune system activity</td>
</tr>
<tr>
<td>rs7574865 (intron)</td>
<td>STAT4</td>
<td>2q32.3</td>
<td>Transcription factor; transduces IL-12, IL-23 and type 1 IFN cytokine signals in T cells and monocytes</td>
</tr>
<tr>
<td>rs2476601 (p.R620W, causal variant)</td>
<td>PTPN22</td>
<td>1p13.2</td>
<td>Intracellular protein tyrosine phosphatase; expressed primarily in lymphoid tissues; involved in TCR signalling</td>
</tr>
<tr>
<td>rs3,733,197 and rs10,516,487 (exon, missense variant)</td>
<td>BANK1</td>
<td>4q24</td>
<td>BANK1 is a B-cell-specific scaffold protein and LYN (MIM #165120) tyrosine kinase substrate promotes tyrosine phosphorylation of inositol 1,4,5-trisphosphate</td>
</tr>
</tbody>
</table>
to be addressed in the investigation of the pathogenesis of SSc.

**Rheumatology key messages**

- SSc belongs to the group of complex genetic disorders.
- The more consistent susceptibility genes identified until now belong to autoimmunity pathways.
- Several genetic factors are shared with other autoimmune diseases.

**Acknowledgements**

**Funding:** This work was funded by Association des Sclérodermiques de France, Institut National de la Santé et de la Recherche Médicale (INSERM), Agence Nationale pour la Recherche (ANR) [Grant no. R07094KS]. It was supported by Groupe Français de Recherche sur la Sclérodermie.

**Disclosure statement:** The authors have declared no conflicts of interest.

**References**

27. Arnett FC, Gourh P, Shete S et al. Major histocompatibility complex (MHC) class II alleles, haplotypes, and epitopes...


