HHS Public Access

Author manuscript

Rheum Dis Clin North Am. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Rheum Dis Clin North Am. 2015 August ; 41(3): 345–366. doi:10.1016/j.rdc.2015.04.001.

Genetics, Epigenetics and Genomics of Systemic Sclerosis

Gloria Salazar, MD and

Division of Rheumatology and Clinical Immunogenetics University of Texas Health Science Center at Houston 6431 Fannin St. MSB 5.270 Houston TX 77030, USA Phone: 713-500-6905 Fax: 713-500-0580 gloria.salazarCintora@uth.tmc.edu

Maureen D. Mayes, MD MPH

Division of Rheumatology and Clinical Immunogenetics University of Texas Health Science Center at Houston 6431 Fannin St. MSB 5.270 Houston TX 77030, USA

Keywords

Genetics; Epigenetics; Systemic Sclerosis; Scleroderma; Gene Expression; microRNA; miRNA; Splicing

1. Introduction

Systemic Sclerosis (SSc) is an autoimmune disease characterized by fibrosis of skin and internal organs, as well as vasculopathy and immune dysregulation. SSc is a clinically heterogeneous disease and presents with 3 distinct subphenotypes, limited, diffuse and “sine” based on the severity of skin involvement. This classification also reflects internal organ involvement which can range from minimal to rapidly progressive disease resulting in premature death. Finally, SSc is characterized by the production of mutually exclusive antinuclear antibodies (ANA) subtypes that are associated with different clinical manifestations, disease phenotypes and prognosis.

SSc is a complex disease that entails abnormalities in several different pathways. Its pathogenesis is not well understood but several studies have established that SSc occurs in a genetically susceptible host presumably after encountering environmental exposures or other external triggers.[1-3] Genetic studies performed so far reveal that multiple genetic loci contribute to disease susceptibility in SSc.[4] The purpose of this review is to discuss the current knowledge of SSc genetics by exploring the observational evidence, the different genetic studies performed to date as well as the most relevant genes discovered by these. We will also explore the concept of gene expression variation and the recently discovered field of epigenetics.

Corresponding Author: Maureen D. Mayes, MD, MPH Phone: 713-500-6905, Fax: 713-500-0580, Maureen.d.mayes@uth.tmc.edu.

Disclosures: The authors have no disclosures.

2. Initial Evidence of the Genetic Influence in SSc

SSc is not inherited in a Mendelian fashion and, although its pathogenesis is unclear, it is believed that gene variants influence not only disease susceptibility but also differences in clinical expression and progression.[2] Here we will discuss some of the initial observational evidence that supported the role of genetics in SSc.

2.1. Ethnic Associations

Differences in prevalence and clinical manifestations among different ethnic groups are evident and support the role of genetics in SSc. Some ethnic groups or subpopulations have higher prevalence of SSc compared to the general population. For example, the Choctaw tribe in Oklahoma is a population with a high prevalence of SSc (two times higher than the expected prevalence) and display a more homogeneous clinical and immunological phenotype than is seen in the general population.[3, 5]

2.2. Shared Genetic Background in Autoimmune Diseases

Several studies have noted that autoimmune diseases (AID) cluster in families (familial aggregation) and that there is co-occurrence in the same individual of two or more AIDs (individual aggregation or overlap syndromes) supporting the concept of a shared autoimmune genetic background among AIDs. In a cross-sectional study including 719 SSc patients, 38% had at least one other overlapping AID. The most frequent overlapping AIDs were autoimmune thyroid disease (AITD) (38%), rheumatoid arthritis (RA) (21%), and Sjogren's syndrome (SS) (18%). Regarding the familial clustering of autoimmunity, 36% of the first-degree relatives had at least one AID of which the most frequent were RA (18%) and AITD (9%). [6] This suggested that AIDs and their manifestations share genetic risk factors and supported the role of genetic influences in SSc.

3. Genetic Study Modalities

The most frequent type of DNA variation consists of a change in an individual nucleotide, known as a single nucleotide polymorphism (SNP). These differences in the DNA sequence may or may not produce functional changes due to modification in the gene expression or alteration in the resulting protein.[7] In general, two basic approaches are used to study genetic associations: the candidate gene approach (CGA) and the genome-wide association study (GWAS). Both methodologies rely on the identification of SNPs and determine the likelihood that the variant occurs more or less frequently in the cases than in the controls (communicated in the form of an odds ratio).[8]

To date, genetic studies have identified SSc susceptibility factors involved in the immune system as well as genes in pathways that play a role in vascular damage and fibrotic processes.[1, 3]

3.1. Candidate Gene Association (CGA)

CGA studies aim to determine specific SNPs associated with disease. Candidate genes are chosen either because they have been associated with other AIDs or because they make biological sense (i.e. are part of pathways hypothesized to be important for the disease

pathogenesis). Therefore the major advantage of the CGA is that one can test for a particular SNP with known functional consequences.[8-10] Using this strategy several novel genetic risk factors for SSc and/or its clinical phenotypes have been identified and these polymorphisms tend to cluster in specific pathways. However, CGA studies are often limited by small cohorts and lack of replication. Genes discovered by CGA are reviewed in tables 1-3.

3.2 Genome Wide Association Studies (GWAS), Meta-GWAS and Pan-Meta-GWAS

GWAS, as opposed to CGA, perform genetic profiling of patients compared to controls by screening SNPs across the entire genome without making *a priori* assumptions about what loci are likely to be involved in the disease. Therefore a major advantage to the GWAS approach is that it is unbiased and can identify novel genes that were not previously suspected to be disease-associated leading to the identification of new pathogenetic mechanisms and yielding many new genetic susceptibility loci.[11, 12] However, a GWAS is limited by the utilized platform as most provide approximately 80% coverage for common polymorphisms in the human genome and miss unusual variants. Another limitation is that GWASs usually investigate SNPs that are in strong linkage disequilibrium (LD) with other SNPs and serve as proxies for them so that the identified SNPs are just a tag for the yet to be identified causal allele.[13]

To date, several GWASs in SSc have confirmed that the major histocompatibility complex (MHC) is the strongest susceptibility loci for SSc. Also multiple non-MHC susceptibility loci have been identified and the most robust associations are in genes related to innate immunity, as well as B- and T-cell activation, which confirmed the concept that SSc is an AID. In 2010 the first robust GWAS in SSc was published that included 2,296 SSc patients and 5,171 healthy controls from the Netherlands, Germany, Spain, and the United States. It revealed that the strongest association was observed at the 6p21 locus corresponding to the MHC. The SNP rs6457617, located in HLA*DQB1, showed the strongest association. In addition, five non-human leukocyte antigen (HLA) loci reached genome-wide significance: TNPO3/IRF5, STAT4, CD247, CDH7, and EXOC2/IRF4 (Tables 1 and 4).[14] Subsequent GWAS with a high resolution marker revealed associations at PSORS1C1 (HLA region), TNIP1, and ras homolog gene family, member B (RHOB) (Tables 2 and 3).[15]

To increase the statistical power to identify low-frequency variants as well as to perform sub-phenotype analysis, it has become popular to merge data from two or more published GWASs and perform a meta-analysis also known as meta-GWAS (MGWAS).[16]

As mentioned before, observational studies support the theory that AIDs share a genetic background. Following this hypothesis, some investigators have merged and analyzed GWASs performed in different AIDs, a concept known as pan-meta-GWAS (PMGWAS). [16] For example, a PMGWAS including SLE and SSc cohorts increased the sample size to a total of 21,109 (6,835 cases) and found, apart from the already known shared susceptibility loci for both diseases, 3 more that were new to SSc (KIAA0319L, PXK and JAZF1, Table 4).[17]

Altogether, these approaches have identified genetic variants T-cell signaling and interferon signaling pathways as associated with SSc susceptibility, and revealed roles in apoptosis, DNA or RNA degradation and autophagy pathways.[10]

3.3. Immunochip

The Immunochip (IC) is a custom SNP genotyping array that provides high-density mapping of AID-associated loci for large cohorts at reduced costs and that specifically targets rare as well as common variants. It can be considered as a hybrid between the CGA and GWAS approaches designed specifically for AIDs as only SNPs that are known to be involved in AIDs are examined. The IC array platform contains 196,524 variants across 186 known autoimmunity risk loci.[13] In 2014, the first SSc IC study was published following in the footsteps of other AIDs. It validated the importance of the HLA region and found new associations including DNASE1L3, the SCHIP1-IL12A locus, and ATG5 as well as a suggestive association between the TREH-DDX6 locus and SSc (Table 3 and 4).[1]

4. Known Genetic Associations with SSc

4.1. Major Histocompatibility Complex (MHC) Region- Human Leukocyte Antigen (HLA) Genes

The strongest genetic association observed in SSc is within the MHC region. The MHC complex located on chromosome 6p21.31 is characterized by the extraordinarily polymorphic HLA alleles and other immunoregulatory genes. Allelic variation in this region has been associated with a wide-range of AIDs including SLE, RA, ankylosing spondylitis (AS), etc.

Polymorphisms in the HLA region have been extensively linked to SSc susceptibility and there are multiple studies that confirm that some HLA types are associated with a general susceptibility to SSc whereas others are more strongly related to particular disease subtypes.[2] Similarly some of these associations are common to several ethnicities whereas others are unique to a specific population group.[18-20]

As previously mentioned, the *HLA-DQB1* gene region conferred the strongest susceptibility for SSc per GWAS. A large multiethnic GWAS published in 2009 revealed that different HLA genotypes not only influence disease susceptibility but are associated with autoantibody expression and vary according to ethnicity. The strongest positive class II associations with SSc in Whites and Hispanics were the *DRB1*1104*, *DQA1*0501* and *DQB1*0301* alleles. In Blacks SSc was associated with *DRB1*0804*, *DQA1*0501* and *DQB1*0301*. *DPB1*1301* showed the highest odds ratio for anti-topoisomerase (ATA) (OR = 14) and anti-centromere (ACA) was associated with *DQB1*0501* and *DQB1*26*.[21]

4.2. MHC region Non-HLA Genes

NOTCH 4 and PSORS1C1 are two genes associated with SSc that are also located in the MHC region but do not code for HLA proteins (Table 2). For example, NOTCH4 encodes a transmembrane protein which plays a role in developmental processes by controlling cell fate decisions and has been implicated in pathways that induce pulmonary fibrosis via TGF-

beta The Notch signaling pathway also controls key functions in vascular smooth muscle and endothelial cells.[22] NOTCH4 polymorphisms have been associated with ACA and ATA positive patients.[22] Little is known about PSORS1C1 other than it is also associated with psoriasis.[23]

4.3. Non-MHC Genes

Multiple non-MHC loci have been associated with SSc. Here we will review the most notable genes discovered to date according to their role in metabolic pathways.

4.3.1. Innate Immunity and IFN Pathway Genes—A growing body of evidence has provided a new paradigm for understanding autoimmunity. Type I interferon (IFN) is now believed to be a central mediator of innate immunity and therefore key for multiple AIDs, SSc included. IFN microarray studies first discovered a key role of the type I IFN in the pathogenesis of SLE and Sjogren's Syndrome (SS) through observation of an IFN "signature" in peripheral blood cells (PBC) and salivary glands. Recent data investigating immunological alterations occurring in early SSc by transcriptional profiling of PBC demonstrated a similar "IFN signature".[24] Type I and II IFN are well known immunomodulators which can also regulate collagen production and therefore are believed to play a key role in the pathogenesis of SSc.[22] Table 1 lists the most notable innate immune genes involved in SSc.

4.3.2. Adaptive Immunity—Associations between defects in B and T cell signaling components and SSc pathogenesis have also been described. T and B cell related gene polymorphisms highlight the importance of these cells and the adaptive immune system in SSc susceptibility. Table 1 lists the adaptive immunity genes and their reported associations with SSc.

4.3.3. Cell Signaling Pathways, Cytokines and Chemokines—Variants in genes that code for several cytokines and cytokine receptors as well as various other signaling pathways have been reported to be associated with SSc. They underscore the importance of the immune system and other non-immune signaling pathways such as cell fate decisions in the pathophysiology of SSc. These genes are summarized in Table 2.

4.3.4. Extracellular Matrix Genes and Others—Several genes involved in extracellular matrix (ECM) deposition have been implicated in the pathogenesis of SSc. Genetic studies in SSc have also revealed the importance of genes and pathways that were previously not considered such as DNASE1L3 (involved in DNA fragmentation) and ATG5 (involved in autophagy) which implicate two novel pathways not previously considered to be important in SSc pathogenesis.[1] Three other novel genes with various functions involved in SSc were discovered by the PWGWAS (Table 3).[17]

4.3.5. The X chromosome—To date, there are three genes associated with SSc susceptibility which are located on the X chromosome, IRAK1, FOXP3 and MECP2.[25, 26] Considering the strong female predilection of SSc, these genes could be of great importance to its pathophysiology. IRAK1 and FOXP3 are vital for T cell function (Table 1).

Interestingly, MECP2 has a key role in DNA methylation and therefore is part of the epigenetic mechanisms for gene regulation (Table 3).[25]

5. Other Sources of Genetic Variation

Very few of the SSc-associated SNPs are found in coding (exonic) regions, but instead are in intronic areas or intragenic regions. It is now clear that there is a complex network of molecular interactions orchestrating gene expression and that these differences are strongly heritable. Genetic mutations that alter any step in this genomic regulatory network can affect gene expression and contribute to regulatory variation. Abnormal gene expression is increasingly recognized as a cause for many disease processes.[27, 28]

5.1. Intronic and Intergenic Mutations

For many years it was believed that the DNA regions that do not code for proteins (known as “non-coding” regions) were not functionally relevant and terms such as “gene deserts” or “junk DNA” were coined for them; however, several studies throughout the years have challenged this idea. GWAS in various diseases (SSc included) have discovered several disease susceptibility loci in these regions and in fact, the vast majority of variants discovered to date are found in non-coding regions (88%).[28-30] It has therefore become increasingly clear that these regions are actually of great importance as they are an integral part of a complicated network that regulates gene expression.

These non-coding regions include two types of DNA sequences: introns and intergenic regions. Introns are nucleotide sequences within a gene that are removed by RNA splicing and eliminated before the final gene product or mature RNA is generated. Intergenic regions are stretches of DNA located between genes. Mutations in both have been discovered in SSc and various other diseases. Mutations in introns can cause a premature stop codon which can generate a truncated (shortened) or abnormally long non-functional or poorly functional protein. Mutations in intergenic regions are believed to affect the normal function of silencer or enhancer sequences. Altering a gene promoter will lead to abnormal gene expression that in turn can cause a disease state.

Multiple polymorphisms of introns located in protein coding genes have been described in SSc and a majority of GWAS hits are actually intronic. Mutations in several intergenic regions have also been described. One example includes the region between SCHIP1 and IL12A that has been strongly associated with SSc, particularly with the limited type and was also identified in celiac disease by IC.[1] Another suggestive association was identified in the 11q23 intergenic region between TREH and DDX6. DDX6 is an RNA helicase that is important for efficient miRNA-induced gene silencing and has been shown to regulate vascular endothelial growth factor (VEGF) under hypoxic conditions, which might provide a clue to the vasculopathy and fibrosis characteristic of this disease, as well as support the importance of epigenetics in SSc.[1] Table 4 lists well known intergenic regions described as susceptibility variants in SSc.

5.2. Splicing Mutations

Recognizing which genomic regions are intronic and which are exonic is key for proper gene expression and this is dependent on a process called “splicing”. Splicing entails the modification of the pre-mRNA transcript by removing introns and joining the exons. The splicing process occurs through a ribonucleoprotein complex known as the spliceosome. This complex is formed by highly dynamic associations and dissociations of hundreds of particles. Many elements can affect the splicing process and understanding how the spliceosome can successfully distinguish exons from introns is vital to decoding gene expression. [31, 32] Splicing mutations can occur and when this happens they can induce exon skipping, form new exon/intron boundaries or activate new cryptic exons. Mutations in introns can also lead to splice mutations. It is estimated that the very high number of currently unclassified genetic alterations may be due to unrecognized gene splicing abnormalities.[31, 32] In SSc we are barely scratching the surface of these intricate mechanisms and their implication for the disease. One study found that an enzyme named Lysyl hydroxylase 2 (LH2), which is vital to the alternatively splicing in collagen biosynthesis might play an important role in SSc. Changes in the pattern of LH2 alternative splicing, can favor increased inclusion of an exon that should be excised thereby increasing the levels of a long transcript which is linked to SSc.[33]

5.3. Expression Quantitative Trait Locus (eQTL)

Given that gene expression levels are considered a heritable and quantitative trait, it is natural to infer that there are associations between gene expression and genotype that can be statistically established and would help identify loci that are associated or linked to gene expression levels. This concept would be especially useful for associations observed in non-coding regions which previously did not have a functional explanation. It was this idea that generated a technique called expression quantitative trait locus (eQTL) mapping which is used to determine the effects of genetic variants on gene expression levels.[27] eQTL studies have established convincing relationships between genetic variants and gene expression by contrasting mean differences in phenotypes among genotypes.[34] The eQTL analyses have been successful in mapping variants to gene expression in several cell types providing a better understanding of the genetics of gene expression, and revealing functional impacts of variants associated with complex traits and diseases.[29]

Gene variants in regulatory regions were classified by eQTL into two types based on the relationship between the genomic location of the mutation and the gene or genes they regulate. A *cis*-regulatory variant alters expression of an associated or nearby gene (local effects). These variants lie either in the basal promoter region located near the transcription start site or in an enhancer located in non-coding sequences surrounding the transcribed region. On the other hand a *trans*-regulatory change regulates expression of genes that are not contiguous and many times are even on different chromosomes (distant effects). eQTL studies show that 25–35% of variants are consistent with *cis*-regulatory effects and the remaining are classified as *trans* acting. Interestingly there is a striking similarity in the amount of *cis*- and *trans* regulatory variation found in genomes from diverse organisms which suggests common gene regulatory mechanisms that were preserved through evolution.[28] To date, eQTL studies have not been performed in SSc but have been

successful for other AIDs such as SLE, making it an attractive future direction in SSc genetics research.[35]

5.4. Epigenetics

Considering that genetics alone is unable to fully account for SSc risk and the polymorphisms that were discovered in non-coding regions, SSc research over the past 5 years has evolved to explore a new field known as epigenetics.[103] Epigenetic variants are defined as the changes in gene function that are inheritable but do not entail a DNA sequence change. Apart from the baseline genomic milieu and the interplay between silencer or enhancer sequences, the epigenetic mechanisms help explain genetic variation given their vital role in gene expression regulation and can account for part of the missed heritability.[10] Some of the epigenetic mechanisms known to date include DNA methylation, histone modification, microRNA (miRNA) and long non-coding RNAs (lnc-RNA) variants. Disruption of any step in this complicated chain of regulatory events can lead to pathology. Here we will briefly discuss our current knowledge in this field.

5.4.1. miRNA and lncRNA—MiRNAs are small, 19-23 nucleotide long, non-coding RNAs that are part of a complicated network which regulates the expression of protein-coding genes. MiRNAs are predicted to regulate up to one third of all human protein-coding genes and most do so by negatively blocking their target mRNA after transcription therefore leading to degradation or repression of translation.[36] MiRNA regulate multiple key pathways that when dysregulated can lead to inflammation, fibrosis, angiogenesis, etc. Various miRNA have been described to be either upregulated or downregulated in tissues of SSc patients compared to healthy controls (HC) including whole skin, fibroblasts or serum. Table 5 depicts the known differentially expressed miRNA in SSc to date. Interestingly some of these miRNA are also dysregulated in other fibrotic diseases such as renal fibrosis or liver cirrhosis. For example, two miRNAs that are shared by multiple fibrotic diseases include the miR-21 and miR-29 families. [37-39]

Only miRNA that have been recapitulated by a second method other than microarray are listed in this Table. Using biocomputational prediction algorithms, which are well established screening tools for miRNA profiling, the miRNAs discovered to date in SSc are usually predicted to target fibrotic pathways such as TGF-beta.[40] Target validation is an important step in miRNA research; it involves proving *in vitro* that a miRNA truly blocks the mRNA that it was predicted to target *in silico*. Table 5 lists the targets that have been validated as opposed to just predicted.

LncRNA as the name depicts are longer RNAs usually >200 nucleotides. LncRNA are also a recently recognized genetic expression regulation mechanism. This field is in its infancy but so far it appears very promising as several lncRNA are reported to regulate immune responses.[41] This is therefore a mechanism that needs to be further investigated in SSc.

5.4.2. Other Epigenetic Mechanisms—DNA methylation occurs when a methyl group is added to CpG dinucleotides that are concentrated in regions called “CpG islands” usually located within promoter regions. This addition promotes a condensed DNA configuration, blocking accessibility to transcriptional activators and thereby inhibiting gene transcription.

Histone modification occurs when histones are acetylated, phosphorylated or methylated influencing the accessibility of chromatin to transcription factors. These mechanisms have been explored to some degree in SSc and are vital for gene regulation expression. Advances in methylation and histone modification in SSc were recently reviewed in great detail by Broen et al.[10]

5.5. Gene-Gene Interactions

The contribution of individual genes to the genetic risk for SSc is modest. Multiple loci are involved and probably interact increasing the risk. It was not until recently that genetic research in the field of SSc started to combine genetic data to determine whether some variants could have an additive risk for SSc susceptibility when found in the same individual. An additional finding that justifies this hypothesis is that many SSc candidate genes map to the same biological pathways. The first successful attempt to study gene-gene interaction in SSc showed that STAT4 (rs7574865) and IRF5 (rs2004640) variants form an additive risk for development of SSc and interstitial lung disease.[42] This analysis was then repeated including the BANK1 polymorphisms and was able to display an additive effect for diffuse SSc susceptibility. In a subsequent analysis, an NLRP1 polymorphism was identified as also interacting with STAT4 and IRF5 contributing to disease risk.[43] These studies underline the importance of gene-gene interaction for development of SSc.

6. Gene Expression Profiling

As mentioned genetics, epigenetics and gene expression go hand in hand. Gene expression profiling (GEP) in SSc has provided insights into the molecular basis of the disease and the underlying changes that occur during disease progression. Global GEP which can be performed with microarrays or high-throughput sequencing allows the simultaneous assessment of thousands of RNA transcripts in a given tissue. It allows examination of gene product interactions along biological pathways.[44] GEP have yielded interesting results that broadened our understanding of SSc. For instance, two studies of early diffuse SSc skin gene expression observed differentially expressed genes associated with increased collagen deposition and ECM synthesis.[45-47] One of them analyzed the GEP of affected and unaffected skin (from skin biopsies) as well as fibroblasts from four dcSSc patients and controls (in cultured cells from skin biopsies). More than 2700 genes were differentially expressed between normal and SSc skin biopsies. Interestingly affected and unaffected skin samples showed nearly identical, disease-specific patterns of gene expression indicating that the genetic expression in SSc is a systemic process irrespective of clinically appreciated abnormalities. The differences in gene expression were mapped to fibroblasts, epithelial, endothelial, smooth muscle, T and B cells.[46] Taking GEP even further, Assassi et al recently observed that 82 skin transcripts distinguished patients with more severe interstitial lung disease. This list included CCL2 (MCP1) which was also associated with SSc by genetic studies (Table 2).[48] The future of GEP is therefore quite promising as there seems to be a link between gene expression in easily accessible tissues such as skin and internal organ involvement which not only offers insight into the systemic process and pathophysiology but could someday aid in risk stratification, prognosis and treatment decisions.

7. Conclusion

The different genetic susceptibility pathways identified so far have offered great insight into the pathophysiology of SSc. The information that has been obtained and that will continue to emerge will provide better disease prognosis, drug responsiveness classification or even predict adverse drug side effects (pharmacogenetics). Discoveries in the field can also lead to identification of novel therapeutic targets and guide drug and biomarker development. As our knowledge of these mechanisms expands, so does our understanding of the intricate pathways necessary to maintain tissue-specific gene expression homeostasis and a healthy state. There is still a big part of the story that remains to be told; the field is evolving daily, offering new insights just around the corner.

Reference List

1. Mayes MD, Bossini-Castillo L, Gorlova O, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 2014; 94(1):47–61. [PubMed: 24387989]
2. Agarwal SK. The genetics of systemic sclerosis. Discov Med. 2010; 10(51):134–143. [PubMed: 20807474]
3. Granel B, Bernard F, Chevillard C. Genetic susceptibility to systemic sclerosis from clinical aspect to genetic factor analyses. Eur J Intern Med. 2009; 20(3):242–252. [PubMed: 19393491]
4. Allanore Y, Dieude P, Boileau C. Genetic background of systemic sclerosis: autoimmune genes take centre stage. Rheumatology (Oxford). 2010; 49(2):203–210. [PubMed: 19923204]
5. Zhou X, Tan FK, Wang N, et al. Genome-wide association study for regions of systemic sclerosis susceptibility in a Choctaw Indian population with high disease prevalence. Arthritis Rheum. 2003; 48(9):2585–2592. [PubMed: 13130478]
6. Hudson M, Rojas-Villarraga A, Coral-Alvarado P, et al. Polyautoimmunity and familial autoimmunity in systemic sclerosis. J Autoimmun. 2008; 31(2):156–159. [PubMed: 18644698]
7. Martin J, Fonseca C. The genetics of scleroderma. Curr Rheumatol Rep. 2011; 13(1):13–20. [PubMed: 20960081]
8. Mayes MD. The genetics of scleroderma: looking into the postgenomic era. Curr Opin Rheumatol. 2012; 24(6):677–684. [PubMed: 23026857]
9. Broen JC, Coenen MJ, Radstake TR. Genetics of systemic sclerosis: an update. Curr Rheumatol Rep. 2012; 14(1):11–21. [PubMed: 22102179]
10. Broen JC, Radstake TR, Rossato M. The role of genetics and epigenetics in the pathogenesis of systemic sclerosis. Nat Rev Rheumatol. 2014; 10(11):671–681. [PubMed: 25136786]
11. Romano E, Manetti M, Guiducci S, Ceccarelli C, Allanore Y, Matucci-Cerinic M. The genetics of systemic sclerosis: an update. Clin Exp Rheumatol. 2011; 29(2 Suppl 65):S75–S86. [PubMed: 21586222]
12. Mayes MD. The genetics of scleroderma: looking into the postgenomic era. Curr Opin Rheumatol. 2012; 24(6):677–684. [PubMed: 23026857]
13. Assassi S, Radstake TR, Mayes MD, Martin J. Genetics of scleroderma: implications for personalized medicine? BMC Med. 2013; 11:9. [PubMed: 23311619]
14. Radstake TR, Gorlova O, Rueda B, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010; 42(5):426–429. [PubMed: 20383147]
15. Allanore Y, Saad M, Dieude P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet. 2011; 7(7):e1002091. [PubMed: 21750679]
16. Martin JE, Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. Hum Genet. 2012; 131(7):1023–1037. [PubMed: 22218928]

17. Martin JE, Assassi S, Diaz-Gallo LM, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum Mol Genet. 2013; 22(19):4021–4029. [PubMed: 23740937]
18. Agarwal SK, Tan FK, Arnett FC. Genetics and genomic studies in scleroderma (systemic sclerosis). Rheum Dis Clin North Am. 2008; 34(1):17–40. [PubMed: 18329530]
19. Tan FK, Arnett FC. Genetic factors in the etiology of systemic sclerosis and Raynaud phenomenon. Curr Opin Rheumatol. 2000; 12(6):511–519. [PubMed: 11092201]
20. Tan FK. Systemic sclerosis: the susceptible host (genetics and environment). Rheum Dis Clin North Am. 2003; 29(2):211–237. [PubMed: 12841292]
21. Arnett FC, Gourh P, Shete S, et al. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. Ann Rheum Dis. 2010; 69(5):822–827. [PubMed: 19596691]
22. Gorlova O, Martin JE, Rueda B, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet. 2011; 7(7):e1002178. [PubMed: 21779181]
23. Reich K, Huffmeier U, Konig IR, et al. TNF polymorphisms in psoriasis: association of psoriatic arthritis with the promoter polymorphism TNF*-857 independent of the PSORS1 risk allele. Arthritis Rheum. 2007; 56(6):2056–2064. [PubMed: 17530646]
24. Wu M, Assassi S. The role of type 1 interferon in systemic sclerosis. Front Immunol. 2013; 4:266. [PubMed: 24046769]
25. Carmona FD, Cenit MC, Diaz-Gallo LM, et al. New insight on the Xq28 association with systemic sclerosis. Ann Rheum Dis. 2013; 72(12):2032–2038. [PubMed: 23444193]
26. D'Amico F, Skarmoutsou E, Marchini M, et al. Genetic polymorphisms of FOXP3 in Italian patients with systemic sclerosis. Immunol Lett. 2013; 152(2):109–113. [PubMed: 23707908]
27. Westra HJ, Franke L. From genome to function by studying eQTLs. Biochim Biophys Acta. 2014; 1842(10):1896–1902. [PubMed: 24798236]
28. Wittkopp PJ. Genomic sources of regulatory variation in cis and in trans. Cell Mol Life Sci. 2005; 62(16):1779–1783. [PubMed: 15968467]
29. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. PLoS Genet. 2014; 10(7):e1004461. [PubMed: 25010687]
30. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106(23):9362–9367. [PubMed: 19474294]
31. De CL, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. 2013; 4(1):49–60. [PubMed: 23044818]
32. Lewandowska MA. The missing puzzle piece: splicing mutations. Int J Clin Exp Pathol. 2013; 6(12):2675–2682. [PubMed: 24294354]
33. Seth P, Walker LC, Yeowell HN. Identification of exonic cis-elements regulating the alternative splicing of scleroderma-associated lysyl hydroxylase 2 mRNA. J Invest Dermatol. 2011; 131(2): 537–539. [PubMed: 20944645]
34. Hulse AM, Cai JJ. Genetic variants contribute to gene expression variability in humans. Genetics. 2013; 193(1):95–108. [PubMed: 23150607]
35. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45(10):1238–1243. [PubMed: 24013639]
36. van RE. The art of microRNA research. Circ Res. 2011; 108(2):219–234. [PubMed: 21252150]
37. Jiang X, Tsitsiou E, Herrick SE, Lindsay MA. MicroRNAs and the regulation of fibrosis. FEBS J. 2010; 277(9):2015–2021. [PubMed: 20412055]
38. He Y, Huang C, Lin X, Li J. MicroRNA-29 family, a crucial therapeutic target for fibrosis diseases. Biochimie. 2013; 95(7):1355–1359. [PubMed: 23542596]
39. Vettori S, Gay S, Distler O. Role of MicroRNAs in Fibrosis. Open Rheumatol J. 2012; 6:130–139. [PubMed: 22802911]

40. Sethupathy P, Megraw M, Hatzigeorgiou AG. A guide through present computational approaches for the identification of mammalian microRNA targets. Nat Methods. 2006; 3(11):881–886. [PubMed: 17060911]
41. Imamura K, Akimitsu N. Long Non-Coding RNAs Involved in Immune Responses. Front Immunol. 2014; 5:573. [PubMed: 25431574]
42. Dieude P, Guedj M, Wipff J, et al. STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis. Arthritis Rheum. 2009; 60(8):2472–2479. [PubMed: 19644887]
43. Dieude P, Wipff J, Guedj M, et al. BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. Arthritis Rheum. 2009; 60(11):3447–3454. [PubMed: 19877059]
44. Assassi S, Mayes MD. What does global gene expression profiling tell us about the pathogenesis of systemic sclerosis? Curr Opin Rheumatol. 2013; 25(6):686–691. [PubMed: 24061076]
45. Sargent JL, Whitfield ML. Capturing the heterogeneity in systemic sclerosis with genome-wide expression profiling. Expert Rev Clin Immunol. 2011; 7(4):463–473. [PubMed: 21790289]
46. Whitfield ML, Finlay DR, Murray JI, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A. 2003; 100(21):12319–12324. [PubMed: 14530402]
47. Gardner H, Shearstone JR, Bandaru R, et al. Gene profiling of scleroderma skin reveals robust signatures of disease that are imperfectly reflected in the transcript profiles of explanted fibroblasts. Arthritis Rheum. 2006; 54(6):1961–1973. [PubMed: 16736506]
48. Assassi S, Wu M, Tan FK, et al. Skin gene expression correlates of severity of interstitial lung disease in systemic sclerosis. Arthritis Rheum. 2013; 65(11):2917–2927. [PubMed: 23897225]
49. Sharif R, Mayes MD, Tan FK, et al. IRF5 polymorphism predicts prognosis in patients with systemic sclerosis. Ann Rheum Dis. 2012; 71(7):1197–1202. [PubMed: 22440820]
50. Dieude P, Guedj M, Wipff J, et al. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. Arthritis Rheum. 2009; 60(1):225–233. [PubMed: 19116937]
51. Dieude P, Dawidowicz K, Guedj M, et al. Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. J Rheumatol. 2010; 37(5):987–992. [PubMed: 20231204]
52. Carmona FD, Martin JE, Beretta L, et al. The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. PLoS One. 2013; 8(1):e54419. [PubMed: 23372721]
53. Ito I, Kawaguchi Y, Kawasaki A, et al. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum. 2009; 60(6):1845–1850. [PubMed: 19479858]
54. Carmona FD, Gutala R, Simeon CP, et al. Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. Ann Rheum Dis. 2012; 71(1):114–119. [PubMed: 21926187]
55. Dieude P, Guedj M, Wipff J, et al. NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis. Ann Rheum Dis. 2011; 70(4):668–674. [PubMed: 21149496]
56. Rueda B, Gourh P, Broen J, et al. BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. Ann Rheum Dis. 2010; 69(4):700–705. [PubMed: 19815934]
57. Gourh P, Agarwal SK, Martin E, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. J Autoimmun. 2010; 34(2):155–162. [PubMed: 19796918]
58. Ito I, Kawaguchi Y, Kawasaki A, et al. Association of the FAM167A-BLK region with systemic sclerosis. Arthritis Rheum. 2010; 62(3):890–895. [PubMed: 20131239]
59. Coustet B, Dieude P, Guedj M, et al. C8orf13-BLK is a genetic risk locus for systemic sclerosis and has additive effects with BANK1: results from a large french cohort and meta-analysis. Arthritis Rheum. 2011; 63(7):2091–2096. [PubMed: 21480188]

60. Dieude P, Bouaziz M, Guedj M, et al. Evidence of the contribution of the X chromosome to systemic sclerosis susceptibility: association with the functional IRAK1 196Phe/532Ser haplotype. Arthritis Rheum. 2011; 63(12):3979–3987. [PubMed: 21898345]
61. Gourh P, Agarwal SK, Divecha D, et al. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis: evidence of possible gene-gene interaction and alterations in Th1/Th2 cytokines. Arthritis Rheum. 2009; 60(12):3794–3806. [PubMed: 19950257]
62. Rueda B, Broen J, Simeon C, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet. 2009; 18(11):2071–2077. [PubMed: 19286670]
63. Tsuchiya N, Kawasaki A, Hasegawa M, et al. Association of STAT4 polymorphism with systemic sclerosis in a Japanese population. Ann Rheum Dis. 2009; 68(8):1375–1376. [PubMed: 19605749]
64. Dieude P, Boileau C, Guedj M, et al. Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. Ann Rheum Dis. 2011; 70(9):1695–1696. [PubMed: 21474487]
65. Dieude P, Guedj M, Wipff J, et al. Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population. Ann Rheum Dis. 2010; 69(11):1958–1964. [PubMed: 20511617]
66. Martin JE, Broen JC, Carmona FD, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Hum Mol Genet. 2012; 21(12):2825–2835. [PubMed: 22407130]
67. Koumakis E, Giraud M, Dieude P, et al. Brief report: candidate gene study in systemic sclerosis identifies a rare and functional variant of the TNFAIP3 locus as a risk factor for polyautoimmunity. Arthritis Rheum. 2012; 64(8):2746–2752. [PubMed: 22488580]
68. Gourh P, Tan FK, Assassi S, et al. Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. Arthritis Rheum. 2006; 54(12):3945–3953. [PubMed: 17133608]
69. Dieude P, Guedj M, Wipff J, et al. The PTPN22 620W allele confers susceptibility to systemic sclerosis: findings of a large case-control study of European Caucasians and a meta-analysis. Arthritis Rheum. 2008; 58(7):2183–2188. [PubMed: 18576360]
70. Diaz-Gallo LM, Gourh P, Broen J, et al. Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Ann Rheum Dis. 2011; 70(3):454–462. [PubMed: 21131644]
71. Gourh P, Arnett FC, Tan FK, et al. Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis. Ann Rheum Dis. 2010; 69(3):550–555. [PubMed: 19778912]
72. Bossini-Castillo L, Broen JC, Simeon CP, et al. A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. Ann Rheum Dis. 2011; 70(4):638–641. [PubMed: 21187296]
73. Coustet B, Bouaziz M, Dieude P, et al. Independent replication and meta analysis of association studies establish TNFSF4 as a susceptibility gene preferentially associated with the subset of anticentromere-positive patients with systemic sclerosis. J Rheumatol. 2012; 39(5):997–1003. [PubMed: 22422496]
74. Dieude P, Guedj M, Truchetet ME, et al. Association of the CD226 Ser(307) variant with systemic sclerosis: evidence of a contribution of costimulation pathways in systemic sclerosis pathogenesis. Arthritis Rheum. 2011; 63(4):1097–1105. [PubMed: 21162102]
75. Bossini-Castillo L, Simeon CP, Beretta L, et al. A multicenter study confirms CD226 gene association with systemic sclerosis-related pulmonary fibrosis. Arthritis Res Ther. 2012; 14(2):R85. [PubMed: 22531499]
76. Mattuzzi S, Barbi S, Carletto A, et al. Association of polymorphisms in the IL1B and IL2 genes with susceptibility and severity of systemic sclerosis. J Rheumatol. 2007; 34(5):997–1004. [PubMed: 17444587]
77. Martin JE, Carmona FD, Broen JC, et al. The autoimmune disease-associated IL2RA locus is involved in the clinical manifestations of systemic sclerosis. Genes Immun. 2012; 13(2):191–196. [PubMed: 22012429]
78. Cenit MC, Simeon CP, Vonk MC, et al. Influence of the IL6 gene in susceptibility to systemic sclerosis. J Rheumatol. 2012; 39(12):2294–2302. [PubMed: 23027890]

79. Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R. Analysis of the 5′ flanking region of the interleukin 10 gene in patients with systemic sclerosis. Rheumatology (Oxford). 2003; 42(11): 1295–1298. [PubMed: 12867583]
80. Hudson LL, Rocca KM, Kuwana M, Pandey JP. Interleukin-10 genotypes are associated with systemic sclerosis and influence disease-associated autoimmune responses. Genes Immun. 2005; 6(3):274–278. [PubMed: 15772682]
81. Bossini-Castillo L, Martin JE, Broen J, et al. A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. Hum Mol Genet. 2012; 21(4):926–933. [PubMed: 22076442]
82. Granel B, Chevillard C, Allanore Y, et al. Evaluation of interleukin 13 polymorphisms in systemic sclerosis. Immunogenetics. 2006; 58(8):693–699. [PubMed: 16832637]
83. Diaz-Gallo LM, Simeon CP, Broen JC, et al. Implication of IL-2/IL-21 region in systemic sclerosis genetic susceptibility. Ann Rheum Dis. 2013; 72(7):1233–1238. [PubMed: 23172754]
84. Agarwal SK, Gourh P, Shete S, et al. Association of interleukin 23 receptor polymorphisms with anti-topoisomerase-I positivity and pulmonary hypertension in systemic sclerosis. J Rheumatol. 2009; 36(12):2715–2723. [PubMed: 19918037]
85. Karrer S, Bosserhoff AK, Weiderer P, et al. The -2518 promotor polymorphism in the MCP-1 gene is associated with systemic sclerosis. J Invest Dermatol. 2005; 124(1):92–98. [PubMed: 15654958]
86. Tan FK, Wang N, Kuwana M, et al. Association of fibrillin 1 single-nucleotide polymorphism haplotypes with systemic sclerosis in Choctaw and Japanese populations. Arthritis Rheum. 2001; 44(4):893–901. [PubMed: 11315929]
87. Wipff J, Giraud M, Sibilia J, et al. Polymorphic markers of the fibrillin-1 gene and systemic sclerosis in European Caucasian patients. J Rheumatol. 2008; 35(4):643–649. [PubMed: 18278837]
88. Fonseca C, Lindahl GE, Ponticos M, et al. A polymorphism in the CTGF promoter region associated with systemic sclerosis. N Engl J Med. 2007; 357(12):1210–1220. [PubMed: 17881752]
89. Zhou X, Tan FK, Reveille JD, et al. Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum. 2002; 46(11):2990–2999. [PubMed: 12428242]
90. Makino K, Jinnin M, Hirano A, et al. The downregulation of microRNA let-7a contributes to the excessive expression of type I collagen in systemic and localized scleroderma. J Immunol. 2013; 190(8):3905–3915. [PubMed: 23509348]
91. Zhu H, Li Y, Qu S, et al. MicroRNA expression abnormalities in limited cutaneous scleroderma and diffuse cutaneous scleroderma. J Clin Immunol. 2012; 32(3):514–522. [PubMed: 22307526]
92. Maurer B, Stanczyk J, Jungel A, et al. MicroRNA-29, a key regulator of collagen expression in systemic sclerosis. Arthritis Rheum. 2010; 62(6):1733–1743. [PubMed: 20201077]
93. Takemoto R, Jinnin M, Wang Z, et al. Hair miR-29a levels are decreased in patients with scleroderma. Exp Dermatol. 2013; 22(12):832–833. [PubMed: 24107002]
94. Tanaka S, Suto A, Ikeda K, et al. Alteration of circulating miRNAs in SSc: miR-30b regulates the expression of PDGF receptor beta. Rheumatology (Oxford). 2013; 52(11):1963–1972. [PubMed: 23893664]
95. Li H, Yang R, Fan X, et al. MicroRNA array analysis of microRNAs related to systemic scleroderma. Rheumatol Int. 2012; 32(2):307–313. [PubMed: 21052672]
96. Nakashima T, Jinnin M, Yamane K, et al. Impaired IL-17 signaling pathway contributes to the increased collagen expression in scleroderma fibroblasts. J Immunol. 2012; 188(8):3573–3583. [PubMed: 22403442]
97. Honda N, Jinnin M, Kira-Etoh T, et al. miR-150 down-regulation contributes to the constitutive type I collagen overexpression in scleroderma dermal fibroblasts via the induction of integrin beta3. Am J Pathol. 2013; 182(1):206–216. [PubMed: 23159943]
98. Honda N, Jinnin M, Kajihara I, et al. TGF-beta-mediated downregulation of microRNA-196a contributes to the constitutive upregulated type I collagen expression in scleroderma dermal fibroblasts. J Immunol. 2012; 188(7):3323–3331. [PubMed: 22379029]

99. Wang Z, Jinnin M, Kudo H, et al. Detection of hair-microRNAs as the novel potent biomarker: evaluation of the usefulness for the diagnosis of scleroderma. J Dermatol Sci. 2013; 72(2):134–141. [PubMed: 23890704]

100. Makino K, Jinnin M, Aoi J, et al. Discoidin domain receptor 2-microRNA 196a-mediated negative feedback against excess type I collagen expression is impaired in scleroderma dermal fibroblasts. J Invest Dermatol. 2013; 133(1):110–119. [PubMed: 22832484]

101. Kajihara I, Jinnin M, Yamane K, et al. Increased accumulation of extracellular thrombospondin-2 due to low degradation activity stimulates type I collagen expression in scleroderma fibroblasts. Am J Pathol. 2012; 180(2):703–714. [PubMed: 22142808]

102. Zhu H, Luo H, Li Y, et al. MicroRNA-21 in scleroderma fibrosis and its function in TGF-beta-regulated fibrosis-related genes expression. J Clin Immunol. 2013; 33(6):1100–1109. [PubMed: 23657402]

103. Sing T, Jinnin M, Yamane K, et al. microRNA-92a expression in the sera and dermal fibroblasts increases in patients with scleroderma. Rheumatology (Oxford). 2012; 51(9):1550–1556. [PubMed: 22661558]

104. Makino K, Jinnin M, Kajihara I, et al. Circulating miR-142-3p levels in patients with systemic sclerosis. Clin Exp Dermatol. 2012; 37(1):34–39. [PubMed: 21883400]

Synopsis

Systemic sclerosis (SSc) is a complex autoimmune disease that occurs in a genetically susceptible host. Genetic studies performed so far reveal that multiple genetic loci contribute to disease susceptibility in SSc. The purpose of this review is to discuss the current knowledge of genetics in SSc by exploring the observational evidence, the different genetic studies and their modalities as well as the most relevant genes discovered by these. The importance of gene expression variation and the different mechanisms that govern it including the recently discovered field of epigenetics are also explored with an emphasis on microRNA (miRNA).

Key Points

- Multiple genes are associated with susceptibility to systemic sclerosis and can lead to alterations in innate and adaptive immunity, cell signaling, extracellular matrix, DNA or RNA degradation and apoptosis or autophagy.
- There are several mechanisms (such as epigenetics and splicing mutations) that influence gene expression and are important for disease in addition to the genetic variants.
- Epigenetic mechanisms govern gene expression at different levels before translation. They are just being explored in SSc and may help explain some of the missing heritability.

Table 1

Adaptive, Innate Immunity and Interferon pathway genes associated with SSc

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
			Innate Immunity and IFN pathways				
IRF5	Interferon Regulatory Factor 5	7q32	rs3757385 rs2004640 rs10954213 rs4728142 rs2280714 rs10488631 rs12537284	GWAS CGA	GSSc dcSSc ATA ILD	Regulates expression of IFN dependent genes	14, 49-53
IRF7	Interferon Regulatory Factor 7	11p15	rs4963128 rs702966 rs1131665	CGA	ACA	Same	54
IRF8	Interferon Regulatory Factor 8	16q24	rs11642873 rs11117425 rs11644034 rs12711490	GWAS	lcSSc ACA	Modulates TLR signaling, regulates IFN gene expression	22
NLRP1	NACHT, LRR and PYD domains-containing protein 1	17p13	rs8182352	CGA	ATA ILD	Interacts with caspases 2 and 9, crucial for the inflammatory response.	55
			Adaptive Immunity				
BANK1	Scaffold protein with Ankyrin Repeat	4q24	rs3733197 rs10516487 rs17266594 rs3733197	CGA	GSSc dcSSc ATA	B-cell-specific scaffold protein, involved in B cell activation	43, 56
BLK	B-Lymphocyte Kinase	8p23	rs13277113 rs2736340 rs13277113	CGA MA	GSSc ACA lcSSc	Downstream BCR signaling	57-59
IRAK1	Interleukin 1 Receptor Associated Kinase-1	Xq28	rs1059702	CGA	GSSc ILD dcSSc ATA	Regulates NF-kB through TCR	25, 60
STAT4	Signal Transducer and Activator of Transcription 4	2q32	rs7574865 rs3821236 rs10168266 rs11889341 rs8179673 rs10181656 rs6752770	CGA GWAS	GSSc lcSSc	Transduces IL12, IL23. Promotes Th1 cells and negatively regulates Th2	14, 15, 42, 61-63
CD247	T-cell receptor zeta (CD3ζ) subunit	1q22	rs2056626	CGA GWAS	GSSc dcSSc lcSSc	Key component of TCR signaling function	14, 15, 64

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
TBX21	T-Box 21 protein (T-bet)	17q21	rs11650354	CGA	GSSc	Critical transcription factor for Th1 development	61
FOXP3	Forkhead Box P3	Xp11	rs3761548 rs2280883	CGA	GSSc lcSSc ACA	Master regulator of Treg cells	26

Abbreviations: GWAS=Genome Wide Association Study, CGA= Candidate Gene Approach, GSSc= Global Systemic Sclerosis, dcSSc=Diffuse Cutaneous Systemic Sclerosis, lcSSc= Limited Cutaneous Systemic Sclerosis, ACA= Anticentromere Antibodies, ATA= Antitopoisomerase (SCL-70) antibodies, ILD= Interstitial Lung Disease, TCR= T-Cell Receptor, BCR= B-Cell Receptor, TLR= Toll Like Receptor, IL-12= Interleukin 12, IL23= Interleukin 23, Th1= T-Helper 1, Th2= T-Helper 2, Treg= T Regulatory Cells, IFN= Interferon, NF-kB= nuclear factor kappa-light-chain-enhancer of activated B cells.

Table 2

Signaling pathways and cytokine genes associated with SSc

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
			Signaling Pathways				
TNFAIP3	Tumor Necrosis Factor α-induced Protein 3	6q23	rs5029939 rs2230926	CGA GWAS	GSSc dcSSc ATA ILD PAH	Negative feedback regulation of the NF-kB pathway	65-67
TNIP1	TNFAIP3-Interacting Protein 1	5q33	rs2233287 rs4958881 rs3792783	GWAS	GSSc	Same	15
PTPN22	Protein Tyrosine Phosphatase Non-Receptor Type 22	1p13	rs2476601	CGA	GSSc ATA	Regulates TCR signal transduction	68-70
TNFSF4	Tumor Necrosis Factor (Ligand) Superfamily, Member 4 (OX40L)	1q25	rs1234314 rs2205960 rs844648 rs4916334 rs10798269 rs12039904 rs8448644 rs844665	CGA	GSSc lcSSc ACA ATA ARA	Binds to OX-40 on T cells (co-stimulatory signal) vital for T cell activation	71-73
CD226	Cluster of Differentiation 226 (DNAX accessory molecule 1)	18q22	rs763361	CGA	GSSc dcSSc ATA ILD	T cell co-stimulatory pathways.	74, 75
SOX5	SRY (Sex Determining Region Y)-Box 5	12p12	rs11047102	MGWAS	ACA	Regulation of embryonic development, determination of cell fate and chondrogenesis.	22
NOTCH4	NOTCH 4	6p21	rs443198 rs9296015	CGA MGWAS	ACA ATA	Controls cell fate decisions.	22
GRB10	Growth Factor Receptor-Bound Protein 10	7p12	rs12540874	CGA MGWAS	lcSSc	Adaptor protein involved in multiple signaling pathways	22
			Cytokines and Chemokines				
IL1B	Interleukin 1β	2q14	rs1143627 rs16944	CGA	GSSc	Crucial for inflammatory responses	76
IL2RA	Interleukin-2 Receptor Subunit Alpha	10p15	rs2104286	CGA	ACA	Treg marker	77

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
IL6	Interleukin 6	7p21	rs2069840	MA	lcSSc	Crucial role in both adaptive and innate immunity	78
IL10	Interleukin 10	1q32	N/A	CGA	GSSc	Anti-inflammatory cytokine	79, 80
IL12RB2	Interleukin-12 Receptor Subunit Beta-2	1p31	rs3790567	GWAS	GSSc	Promotes Th differentiation into Th1 cells	81
IL13	Interleukin 13	5q31	rs2243204	CGA	GSSc	Secreted by activated Th2 cells shown to be involved in fibrosis	82
IL21	Interleukin 21	4q27	rs6822844 rs907715	CGA	GSSc	Potent immunomodulatory cytokine	83
IL23R	Interleukin 23 Receptor	1p31	rs11209026 rs11465804	CGA	ATA	Stabilize the Th17 phenotype	84
MCP1 (CCL2)	Monocyte Chemotactic Protein 1 (CCL2)	17q12	N/A	CGA	GSSc	Recruits monocytes, memory T cells and dendritic cells to sites of inflammation	85

Abbreviations: GWAS=Genome Wide Association Study, MGWAS= Meta-GWAS, MA= Meta-Analysis, CGA= Candidate Gene Approach, GSSc= Global Systemic Sclerosis, dcSSc=Diffuse Cutaneous Systemic Sclerosis, lcSSc= Limited Cutaneous Systemic Sclerosis, ACA= Anticentromere Antibodies, ATA= Antitopoisomerase (SCL-70) antibodies, ARA= Anti-RNA Polymerase-III, ILD= Interstitial Lung Disease, PAH= Pulmonary Arterial Hypertension, TCR= T-Cell Receptor, Th1= T-Helper 1, Th2= T-Helper 2, Th17= T-Helper 17, Treg= T Regulatory Cells, NF-kB= nuclear factor kappa-light-chain-enhancer of activated B cells.

Table 3

Extracellular matrix and other genes associated with SSc

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
			Extracellular Matrix				
FBN1	Fibrillin 1	15q21	N/A	CGA	GSSc	Glycoprotein, structural component of microfibrils	86, 87
CSK	C-src Tyrosine Kinase	15q24	rs1378942	GWAS	GSSc	Inactivates the c-Src kinases that lead to fibrosis	66
CTGF (CCN2)	Connective Tissue Growth Factor	6q23	rs6918698	CGA	GSSc	Promotes fibroblast proliferation and ECM production.	88
SPARC (Osteonectin)	Secreted Protein Acidic and Rich in Cysteine	5q33	N/A	CGA GWAS	GSSc	Modulates cell-matrix interactions	89
			Other Genes				
DNASE1L3	Deoxyribonuclease 1-like 3	3p14	rs35677470	IC	ACA	DNA fragmentation during apoptosis	1
ATG5	Autophagy Protein 5	6q21	rs9373839	IC	GSSc	Protein necessary for autophagy	1
RHOB	Ras Homolog Family, protein B	2p24	rs13021401 rs342070	GWAS	GSSc	Cell migration	15
PXK	PX Serine Threonine Kinase	3P14	rs2176082	PMGWAS	ACA	Trafficking of the EGFR	17
JAZF1	Juxtaposed with Another Zinc Finger Protein 1	7p15	rs1635852	PMGWAS	GSSc	Associated with skeletal frame size and height, likely related to bone morphogenesis and collagen deposition	17
KIAA0319L	KIAA0319-Like protein	1p34	rs2275247	PMGWAS	lcSSc	Unknown	17
MECP2	Methyl-CpG Binding Protein 2	Xq28	rs17435	CGA	GSSc dcSSc	Binds to methylated DNA, transcriptional regulation of methylated genes	25
PSORS1C1	Psoriasis Susceptibility 1 Candidate 1	6p21	rs3130573	GWAS	GSSc	Unknown	15

Abbreviations: GWAS=Genome Wide Association Study, PMGWAS= Pan-Meta-GWAS, CGA= Candidate Gene Approach, IC= Immunochip, GSSc= Global Systemic Sclerosis, dcSSc=Diffuse Cutaneous Systemic Sclerosis, lcSSc= Limited Cutaneous Systemic Sclerosis, ACA= Anticentromere Antibodies, ECM= Extracellular Matrix, EGFR= Epidermal Growth Factor Receptor, NF-kB= nuclear factor kappa-light-chain-enhancer of activated B cells.

Table 4

Intergenic regions associated with SSc

Gene	Product	Location	Polymorphisms	Approach	Phenotype	Function	References
SCHIP-IL12A	N/A	3q25	rs77583790	IC	GSSc lcSSc ACA	Unknown	1
TREH-DDX6	N/A	11q23	rs7130875	IC	GSSc	Unknown	1
TNPO3-IRF5	N/A	7q32	rs12537284	GWAS	GSSc	Unknown	14
CDH7	N/A	18q22	rs10515998	GWAS	GSSc	Unknown	14
EXOC2-IRF4	N/A	6p25	rs4959270	GWAS	GSSc	Unknown	14

Abbreviations: GWAS=Genome Wide Association Study, IC= Immunochip, GSSc= Global Systemic Sclerosis, lcSSc= Limited Cutaneous Systemic Sclerosis, ACA= Anticentromere Antibodies.

Table 5

Dysregulated miRNA in Systemic Sclerosis

miRNA	Location	Predicted Targets	Validated	Induced or Suppressed by	Tissue	Method	Phenotype	References
				Downregulated				
let-7a	9q22	COL1A1 COL1A2	Yes	Unknown	Skin Serum Fibroblasts	TPCR PCRa	GSSc	90
miR-29a	7q32	COL1A1 COL1A2 COL3A1	Yes	TGF-β PDGF-B IL-4	Skin Serum Hair	TPCR	GSSc	91-93
miR-30b	8q24	PDGFR-β	Yes	TGF-β	Serum	TPCR	GSSc dcSSc	94
miR-125b	11q24	SMAD5 CD28	No	Unknown	Skin	mirA TPCR	GSSc	91, 95
miR-129-5p	7q32	COL1A1	Yes	Unknown	Fibroblasts	PCRa TPCR	GSSc	96
miR-145	5q32	SMAD3	No	Unknown	Skin Fibroblasts	TPCR mirA	GSSc	91, 95
miR-150	19q13	Integrin β3	Yes	Unknown	Skin Serum Fibroblasts	TPCR PCRa ISH	GSSc	97
miR-196a	17q21	COL1A1 COL1A2	No	TGF-β	Hair Skin Serum Fibroblasts	TPCR PCRa ISH	GSSc dcSSc P&U	98-100
miR-206	6p12	TGFB1 TGFB2 SMAD5 Integrin α2	No	Unknown	Skin	mirA TPCR	GSSc	91, 95
				Upregulated				
let-7g	3p21	COL1A2 COL2A1 COL5A2 TGFB2R	No	Unknown	Skin	mirA TPCR	GSSc	95
miR-7	9q21	COL1A2	No	TSP-2	Skin Fibroblasts	PCRa ISH	GSSc	101
miR-21	17q23	SMAD7	Yes	TGF-β	Skin Fibroblasts	TPCR mirA	GSSc	91, 102
miR-92a	13q31	MMP1	No	TGF-β	Serum Fibroblasts	TPCR	GSSc <TA	103

miRNA	Location	Predicted Targets	Validated	Induced or Suppressed by	Tissue	Method	Phenotype	References
miR-142-3p	17q22	Integrin αV	No	TGF-β	Serum	TPCR	GSSc	104

Abbreviations: GSSc= Global Systemic Sclerosis, dcSSc= Diffuse Cutaneous Systemic Sclerosis, COL1A1= Collagen 1 alpha 1, COL1A2= Collagen 1 alpha 2, COL3A1= collagen 3 alpha 1, PDGFR-β= Platelet Derived Growth Factor Receptor Beta, PDGF-B= Platelet Derived Growth Factor Beta, TGFB2R=Transforming Growth Factor Beta Receptor II, TGF-β= Transforming Growth Factor Beta, TGFB1= TGF Beta 1, MMP1= Matrix Metalloproteinase 1, IL-4= Interleukin 4, TSP-2= Thrombospondin 2, TPCR= Targeted Polymerase Chain Reaction, PCRa=PCR array, mirA= miRNA Array, ISH= In-Situ Hybridization, P&U=Pits and Ulcers, TA= Telangiectasias.