Molecular “omic” signatures in systemic sclerosis

Bhaven K. Mehta, Monica E. Espinoza, Monique Hinchcliff, Michael L. Whitfield

Abstract

Systemic sclerosis (SSc) is a connective tissue disorder characterized by immunologic, vascular, and extracellular matrix abnormalities. Variation in the proportion and/or timing of activation in the deregulated molecular pathways that underlie SSc may explain the observed clinical heterogeneity in terms of disease phenotype and treatment response. In recent years, SSc research has generated massive amounts of “omics” level data. In this review, we discuss the body of “omics” level work in SSc and how each layer provides unique insight to our understanding of SSc. We posit that effective integration of genomic, transcriptomic, metagenomic, and epigenomic data is an important step toward precision medicine and is vital to the identification of effective therapeutic options for patients with SSc.

Keywords: Scleroderma, systemic sclerosis genomics, computational biology, transcriptomics, epigenomics, metagenomics

Introduction

Systemic sclerosis (SSc) is a rare, autoimmune disease of unknown etiology that is characterized by vascular abnormalities, immunologic aberrancies, and extracellular matrix (ECM) deposition in the skin and internal organs (1). Like many autoimmune disorders, women are disproportionately affected (the ratio female:male is >3:1) (2). Patients are typically diagnosed in their fifth and/or sixth decade of life. On the basis of the 1980 American Rheumatism Association primary criteria, varying prevalence and incidence of SSc by geographic locations have been identified, with low estimates from northern Europe and Japan (prevalence: <150 per million, incidence: <10 per million per year) and higher estimates from the United States, Canada, southern Europe, and Australia (prevalence: >150 per million, incidence: >10 per million per year) (3). A study assessing the prevalence of various rheumatic conditions found that SSc affects approximately 49,000 individuals in the United States (4). However, with updated and more sensitive American College of Rheumatology-European League against Rheumatism 2013 criteria (5), current prevalence and incidence rates of SSc are likely higher.

Patient heterogeneity is an intrinsic feature of SSc and manifests at the phenotypic level through extent of skin fibrosis, autoantibody status, organ involvement, and temporal and spatial progression and manifests at the molecular level where transcriptomic studies have identified subsets of patients with vastly differing disease signatures (6-9). Two clinical subtypes defined by the extent of skin involvement—diffuse cutaneous SSc (dcSSc) and limited cutaneous SSc (lcSSc)—have been described. The dermal fibrosis in patients with lcSSc is distal to their knees and elbows and may involve the face, whereas patients with dcSSc will also have fibrosis proximal to the elbow or knees and possibly involve their trunk. In general, patients in each subtype display different disease trajectories and serum autoantibodies (1). For instance, patients with lcSSc is more likely to express anticentromere serum autoantibodies and to develop pulmonary arterial hypertension, whereas patients with dcSSc are more likely to demonstrate anti-topoisomerase I or anti-RNA polymerase III serum autoantibodies and are at increased risk for early onset renal, cardiac, and pulmonary complications (10). This heterogeneity and inability to predict disease course on a per-patient basis has prompted the search for different methods to further define the patient population using more granular data.

In studies of breast cancer, intrinsic subsets based on distinct gene expression signatures have been identified and have been found to be associated with differences in prognosis and treatment response (11). In 2013, the Food and Drug Administration approved the Prosigna PAM50 Breast Cancer risk of recurrence assay, a genomic test to assess the activity of a targeted set of genes in patients with hormone-receptor–positive breast cancer (12, 13). Estimating the risk of recurrence helps physicians determine the appropriate hormonal therapy duration. Similarly, molecular SSc subsets (i.e., fibroproliferative, inflammatory, limited,
and normal-like) that are distinct from the clinically observed subsets (i.e., lcSSc and dcSSc) have been identified by comparing gene expression in skin and esophageal biopsies from patients with SSC with expression in healthy individuals (6, 14-17). Interestingly, analyses of skin gene expression data from patients that did not meet primary endpoints in clinical trials of patients with SSC on abatacept (an inhibitor of T-cell activation) show that patients in the inflammatory intrinsic subset are most likely to demonstrate improvement in skin disease during treatment (18, 19).

Characterizing the heterogeneity of SSC using transcriptomics

The systemic nature of SSC has been well-documented clinically; however, manifestations at the molecular level have been more difficult to investigate because they require patient biopsies. The increasing availability of "omic" approaches coupled with the decreasing cost of sequencing and increasing access to tissue biopsies has enabled researchers to interrogate tissues from patients with SSC to better understand disease heterogeneity.

The systemic manifestations of SSC were first assessed at the transcriptomic level by performing DNA microarrays on RNA isolated from biopsies of clinically involved forearm skin (lesional skin) and clinically uninvolved back skin (non-lesional skin) (20). The first result was clear that SSC skin biopsies were distinct from healthy control skin biopsies. Moreover, non-lesional skin from patients with SSC was more similar to lesional SSC skin and showed similar patterns of differential gene expression compared with healthy control skin. These results indicated that gene expression was altered systemically, regardless of clinically observed fibrosis. Another study compared the gene expression of cultured dermal fibroblasts from skin biopsies of monozygotic and dizygotic twins, the majority of which were discordant for SSC (i.e., one with SSC and the other healthy), to healthy control fibroblasts of unrelated individuals (21). The authors found similar gene expression between monozygotic twins where at least one individual was affected, suggesting that there is a strong genetic predisposition to SSC beyond what can be clinically observed (21).

Although early studies using whole skin biopsies or cultured dermal fibroblasts focused on differences between SSC and healthy controls, more recent work has sought to unravel the heterogeneity that exists within the patient population with SSC. The SSC intrinsic subset classification was first described after DNA microarrays were applied in a study using RNA isolated from lesional forearm and non-lesional back skin biopsies (6). The SSC intrinsic subsets, aforementioned, are labeled inflammatory, fibroproliferative, normal-like, and limited (Figure 1). The inflammatory subset is associated with an active immune response and defense response; the fibroproliferative subset is associated with proliferation and cell cycle programs; and the normal-like subset is associated with a distinct lack of inflammatory signature coupled with fatty-acid metabolism. The limited subset is the least well-characterized but shows deregulation of pathways associated with cell adhesion, cardiovascular system development, ECM, and immune and inflammatory responses (22). These SSC intrinsic subsets have been shown to be relatively stable throughout disease course and unlikely to change over time without an intervention (7).

Although subsets were first identified through DNA microarray analysis of skin (6), they have since been validated in additional SSC cohorts (7-9), in other affected organs (23), and on different platforms (24). The patient heterogeneity defined by the intrinsic molecular subsets may reflect the complex polygenic underpin-
The heterogeneity in the patient population with SSc has rendered identification of animal models and in vitro models difficult. In fact, no single in vitro or in vivo model fully recapitulates all facets of the human disease. Gene expression analyses of tissue from various mouse models of scleroderma have enabled identification of models that can be used to study particular SSc intrinsic subsets (25). Specifically, a study found that the scleroderma graft-versus-host disease and bleomycin-induced fibrosis mouse models have changes in gene expression similar to those found in patients in the inflammatory subset (12, 25). In contrast, Tsk2/+ mice were more similar to the fibroproliferative subset of patients (25, 26). Since the first description of the SSc molecular intrinsic subsets, our understanding of the molecular landscape and inter-patient heterogeneity has vastly increased and many have applied novel computational frameworks to further study the intricacies of SSc. Recent work utilized consensus clustering, a method that identifies sets of genes that are co-expressed in a similar manner across multiple datasets, to identify conserved sets of genes across three cohorts of skin of patient with SSc. This work defined genes (PLAUR, RAC2, LCP, CXCR4, CD14, THY1) that bridged inflammatory and proliferative processes (27) and suggested that specific single nucleotide polymorphisms (SNPs) were playing key roles in the regulation of abnormal gene expression in SSc. Important ly, this work further implicated the alternatively activated M2-like macrophage as an integral component in SSc biology. Activated macrophages may play an important role in disease pathogenesis and progression in SSc because they regulate vascularization and inflammation and secrete pro-fibrotic mediators, including transforming growth factor-beta (TGF-β), interleukin 6 (IL-6), and C-C motif chemokine ligand 2. The role of activated M2-like macrophages in SSc has been extensively described in a recent review (28).

A meta-analysis of 10 datasets with 573 samples from 321 patients with SSc was conducted to identify deregulated molecular pathways that are consistently observed across multiple end-target organs. Data from skin, esophagus, lung, and blood from patients with SSc were analyzed and again pointed to a deregulated immune-fibrotic axis (29). Importantly, this work identified inflammatory modules (sets of genes associated with immune processes) in the skin that were associated with severe phenotypes (e.g., pulmonary fibrosis) and implicated M2/alternatively activated macrophages as a key component of the immune-fibrotic axis. Although this work found that subset-like signatures were conserved in end-target organs, it is unknown whether the SSc tissue gene expression signature is conserved within an individual or whether distinct SSc signatures are present in different involved organs.

In SSc, the complexity that exists at the molecular level has, to date, been largely studied using bulk sequencing, where the average gene expression of all cells in a biopsy is used as the representative value for that biopsy. Single-cell sequencing (single-cell RNA-Seq), whereby gene expression is determined for each distinct cell population, represents a new era of transcriptomic research. In recent systemic lupus erythematosus work, investigators performed single-cell RNA-Seq on circulating monocytes and identified distinct gene expression signatures between classical (CD14++/CD16+) and non-classical monocytes (CD14−/CD16−) (30). The authors also found monocyte sub-populations via unsupervised hierarchical clustering that were related to type 1 interferon and disease activity (30). Single-cell RNA-Seq provides more granular data that can be used to answer many questions in SSc, including the question of cell-type specific contributions and the interplay between the immune system and fibrotic tissues. A small SSc study (1 patient with SSc and 1 control) performed single-cell RNA-Seq on forearm skin endothelial cells and re-identified a marker of vascular injury (THBS1 gene/thrombospondin) that had previously been described as important in SSc while providing new candidate genes (APLBR gene/apelin receptor and HSPG2 gene/heparan sulfate proteoglycan) that may warrant further study (31). Studies like this in SSc may provide a clearer picture of the deregulation occurring in the context of individual cell types and allow signals to become more prominent that would have previously been generalized under summary measures such as median or average expression.

Molecular signatures in therapeutic trials for SSc

Although the intrinsic subsets communicate the heterogeneity of the patient population with SSc, they remain largely absent from informing clinical trial enrollment criteria. However, it is possible that drugs abandoned as ineffective in SSc were effective in a particular subset of patients. Retrospective analyses of negative clinical trials have identified groups of patients in whom a particular drug may have been effective, and this information could possibly be used to inform the design of future larger studies (32, 33). As expected, intrinsic subset designation seems to play an important role in response to therapy (34). For instance, patients with SSc classified in the inflammatory intrinsic subset are more likely to respond to mycophenolate mofetil (8) and abatacept (19), whereas patients belonging to the fibroproliferative intrinsic subset may be more likely to benefit from imatinib mesylate (35) or dasatinib (36). Some work has also identified patient populations that worsened on therapy, such as inflammatory patients on dasatinib (36)—an equally important consideration when selecting therapy. Importantly, although the intrinsic subset classification is a useful way to divide patients with SSc, some work has shown that specific pathways that are relevant to a drug’s mechanism of action should be directly assessed. One such case may be the requirement of high TGF-β receptor type 1/platelet derived growth factor receptor beta-signaling for nilotinib improvement (37). In a step toward precision medicine, future Phase 2 clinical trials in SSc should focus on identifying patient subsets that respond to treatment on a molecular basis so that Phase 3 studies can be enriched for this set of patients. Our understanding of SSc supports the notion that uncovering the disease biology and mechanism in an individual patient through the measurement of gene expression is critical for the development of effective treatments in SSc.

Studies of SSc genetic risk

Although SSc etiology is poorly understood, as with many autoimmune disorders, a set of complex genetic and environmental factors have been implicated in the disease. Twin studies have revealed low concordance for disease; however, the presence of antinuclear antibodies were significantly more concordant between monozygotic twins (40), suggesting that inherited genetics may predispose to the SSc development but that a second insult may be required for pathogenesis to begin. Genome-wide association studies (GWAS) in many autoimmune disorders have identified similar putative causal mutations in immune-related gene loci (STAT4, IRF5, CD247, and a number
of different HLA genes) in the Caucasian population (41). Thus, the contribution of individual genetic changes in SSC is likely impacted by a strong environmental component.

Early candidate–gene approaches used a priori knowledge of genes purported to be involved in SSC and investigated differences in genomic variants between patients with SSC and healthy controls. These candidate–gene studies implicated mutations in TGFβ1 (42, 43), IL-1A (44), SPARC (45), PTPN22 (46), ILRA1 (47), CTGF (48), IL23R (49), IRF5 (50), STAT4 (51), FAS (52), TBX21 (53), BANK1 (54), FAM167A-BLK (55), TNFSF4 (41), KCNA5 (56), IRAK1 (57), and IRF7 (58) with SSC development. The advent of SNP chips allowed GWAS to interrogate the genome more comprehensively and in an unbiased manner. Five GWAS (41, 59-62) and three follow-up GWAS (63-65) in patients with SSC and controls have identified many HLA (HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DBP1) and non-HLA associations of varying significance with STAT4, CD247, and IRF5 frequently associated with SSC (66).

Analyses continue to identify new loci but also to investigate specific genetic loci variants that GWAS has identified as important, such as NOTCH4 (67), GSDMA (68), PRDM1(68), and IRF4 (69). Importantly, these studies all contribute to the growing body of literature regarding the complex genetic make-up of autoimmune disease. For example, a rare NOTCH4 variant was identified via whole-exome sequencing (WES) of an 8-year old with dcSSc and a family history of SSC (grandfather and maternal aunt) (70). The authors concluded that this NOTCH4 mutation could be a rare, but penetrant, SSC risk variant that may underlie the NOTCH4 GWAS signal and contribute to SSC risk.

WES studies have identified interesting variants at a higher resolution; however, owing to the heterogeneity in patient populations, conclusions are limited to the population studied. For example, in 2016, Gao et al. (71) identified that the gene variant ATP8B4 was enriched in European-Americans with SSC; however, another study of European patients was unable to validate this variant (72). Mak et al. (73) studied 32 patients with dcSSc and interstitial lung fibrosis (23 had >90% European ancestry) and 17 healthy controls using WES and identified 70 gene variants (COL4A3, COL4A4, COL5A2, COL13A1, COL22A1) that were significantly enriched in ECM-related pathways (p=0.002).

Figure 2. Model of the gene-gene network constructed from multiple transcriptomic studies of SSC skin. This illustration identifies key pathways of disease in SSC and how specific polymorphisms link disease pathways. Purple represents processes associated with the inflammatory subset. Red represents processes associated with the fibroproliferative subset. Adapted with permission from Mahoney et al. (27).

Adaptive Immunity

Interferon

ECM

M2 macrophages

Proliferation

GRB10

LCP2

CXCR4

RAC1

PLAUR

CD14

THY1

IRAK-I

PAK

ERK/MAPK

N. H. Yee et al. Eur J Rheumatol 2020; 7(Suppl 3): S173-80

A recent gene-level association analysis of SSC revealed four novel loci for white patients—STAT1, FGR2C, NIPSNAP3b, and SCT (74). An exploratory analysis comparing gene-level analysis results between African American (AA) and white patients demonstrated mutations TNFAIP3, FGR2C, PINX1, UNC5C, and CLEC16A at varying significance levels and susceptibility loci SCN4B, SCN2B, MRPL28, TME8, NME4, and DEC2R to only be significantly found in AA individuals. It is of note that in this study, the term white was used rather than Caucasian; these delineations should be recognized in the interest of questions regarding the ractioethnic diversity of individuals with SSC. In another study, the CD247 locus associated with Caucasian patients with SSC was not found to be associated in a population of Han Chinese patients with SSC (75).

Although the identification of genetic loci associated with SSC development may start to reveal a clearer picture for some families (NOTCH4 case study), the genetic causes in under-sampled populations remain elusive. These discrepancies highlight that more ractioethnically expansive research is needed to understand disease heterogeneity among differing genetic backgrounds as with most autoimmune diseases. To begin to address these concerns, the Genome Research in African American Scleroderma Patients (GRASP) consortium has specifically focused on comprehensive genetic studies in AA patients with SSC. Interim GRASP study results found an ATP8B4 variant that was associated with SSC in patients who self-identified as being of European ancestry (71) but was not present in patients who self-identified as AA. Significant enrichment in mutations in ECM-related pathways (76) was found in both self-identified populations. Meta-analyses across varying populations may begin to highlight key loci (e.g., TNFAIP3 and ECM-related pathways); however, to begin to unravel the intricacies of SSC racialoethnic differences must be considered. It is entirely plausible that patients of differing ractioethnic backgrounds arrive at the disease state by differing mechanisms and may require distinct medical intervention.

In silico integration of our knowledge of SNPs and transcriptomic studies provides vital hypotheses that can be further tested at the bench. One such study generated a gene-gene network using network analysis on consensus clusters derived from three distinct datasets, highlighting the molecular subsets and enriching with previously associated risk loci in the context of the intrinsic molecular subsets (77). The analysis clearly identified functional modules related to interferon activation, alternatively activated macrophages, adaptive immunity, ECM remodeling, and cell proliferation. The work showed that many risk loci, while each only having modest odds ratios, likely underlie the changes in gene expression that lead to intrinsic molecular subsets in SSC. The authors proposed a model of interaction between modules of gene expression, placing key loci such as CXCR4 and THY1 as important bridges between key components of SSC biology (Figure 2).

Epigenetics in SSC

SSC has genetic underpinnings but an environmental insult is also thought to be associated
with the disease onset. Epigenetic modifications mediate the interaction between environmental exposure and genetic code and are heritable. As such, they can be regarded as an additional level of regulation to be considered for therapeutic targeting.

DNA methylation of cytosine residues is one of the most-studied epigenetic modifications where a methyl group is added to a cytosine to form 5-methylcytosine by DNA methyltransferases (DNA (cytosine-5)-methyltransferase 1 [DNMT1], DNMT3a, and DNMT3b). In humans, this methylation of cytosines occurs more often at CpG dinucleotides and, if found in gene promoters, the methylation leads to repression of transcription by recruitment of methyl-CpG-binding domain proteins (MBDs). With enough density, MBDs can further recruit histone deacetylases (HDACs), causing further DNA compaction resulting in strong repression of a region. Thus, changes in methylation status can alter gene expression.

As with genetic studies, early epigenetic work in SSc required a targeted strategy. In 2006, a study demonstrated that treatment with DNMTs and HDAC inhibitors (trichostatin A [TSA]) resulted in normalization of collagen levels in cultured dermal fibroblasts (78). This repression of collagen production was linked to the gene FLI1; specifically, treatment with epigenetic inhibitors reversed the high methylation present on the FLI1 promoter in SSc fibroblasts and biopsies, resulting in normalized collagen levels. This study was the first to demonstrate a link between epigenetic changes and fibrotic potential in SSc cultured dermal fibroblasts. In 2007, a study used the same HDAC inhibitor (TSA) in dermal fibroblasts and showed that the inhibitor resulted in decreased cytokine-induced transcription of ColA1 and upregulation of the cell cycle inhibitor p21, resulting in reduced fibroblast proliferation (79). Another study found that global methylation in circulating CD4+ T cells in patients with SSc was significantly reduced relative to healthy controls, concurrent with reduced levels of DNMT1, MBD3, and MBD4 (80). Follow-up work found that CD4+ T cells were further implicated in SSc pathogenesis because the overexpression of CD70 (81) and CD11a (82) (T-cell surface markers) caused by hypomethylation was associated with SSc. In the subcutaneous bleomycin-induced fibrosis model, treatment with TSA resulted in less dermal ECM deposition (79). Each of these studies implicated epigenetic regulation as a factor in fibrotic disease.

Altorok et al. (83) performed a genome-wide DNA methylation analysis on dermal fibroblasts from patients with lcSSc and patients with dcSSc and compared them with controls. The study found 2,710 (dcSSc) and 1,021 (lcSSc) differentially methylated regions with an overlap of 203 (118 hypomethylated and 6 hypermethylated) CpG sites. It should be noted that the hypomethylated genes included those relevant to SSc in immune and fibrotic function (RUNX1, RUNX2, RUNX3, ADAM12, COL23A1, COL4A2, and MYO1E). As with the genetic and transcriptomic studies, Altorok et al. (83) showed that although commonalities exist (203 overlap) within the population with SSc, vast differences are evident underscoring the need for additional studies. An interesting direction would be to interrogate the epigenetic landscape present in multiple organs of single individuals and to investigate whether these differences in the heterogeneous patient populations are systemically conserved within a single individual across organs.

Extracting information on the microbiome from SSc “omic” signatures

As with genetic and epigenetic contributions, little is known about the environmental insults that could trigger or exacerbate the genetic predisposition in an individual to result in disease. Common factors that have been implicated in other autoimmune disorders, such as cigarette smoking and alcohol consumption, are not recognized risk factors for SSc development (84, 85), whereas exposure to occupational hazards such as silica dust and solvents has been repeatedly associated with the SSc risk (86).

An important consideration is the microbiome that constitutes an individual’s most immediate environment. The microbiome has been characterized in a variety of SSc-affected tissues using different approaches. Targeted in situ hybridization experiments studies of the microbiome identified elevated Epstein-Barr Virus non-coding RNA in skin samples (87). The 16S sequencing studies of the lower gastrointestinal tract through colonic lavage and via the collection of stool specimens have found a decrease in commensal bacteria and an increase in potentially pathobiont bacteria in patients with SSc compared with controls. Notably, an increase in the commensal Lactobacillus was found, which could potentially play a role in altered SSc gastrointestinal motility (88).

The microbiome has been indirectly assessed through metagenomics, where RNA-Seq of patient epithelial tissue samples can yield information on the microbial communities that reside in these tissues. Host reads can be filtered out of the total read set, leaving exogenous mi-crobial reads, which can be identified using Basic Local Alignment Search Tool against a universal database for taxonomic characterization (89). Here, the microbial community, which interacts and evolves with its host through health and disease, is captured alongside the gene expression of the patients’ tissue, uncovering a previously hidden dimension of data. This approach identified an overabundance of Rhodotorula glutinis on lesional forearm skin of patients with SSc (89). A second follow-up study of SSc skin using the same approach, but now focused on later stage disease, did not find R. glutinis but rather a decrease in taxon such as Propionibacterium and an increase in genera such as Bunkholderia, Citrobacter, and Vibrio, with dysbiosis largely associated with inflammatory gene signatures and disease duration (90).

These studies suggest that a maladapted microbiome is associated with SSc and includes differing species richness and relative abundance of commensal microbes in affected tissue, as determined through targeted and untargeted studies of the microbiome by methods including metagenomics. It is important to note that the associations with the microbiome, to date, are unable to determine causality.

Conclusion

“Omics” level interrogation of tissue from patients with SSc has risen steeply in recent years. Genomic, transcriptomic, epigenetic, and metagenomic data are now available for study in addition to cytokine and protein expression and cell-type information to elucidate more of SSc disease biology. Newer studies that layer the information from many ‘omics’ approaches in the same patient tissue sample move investigations toward the application of precision medicine (91), which we propose should be the direction of future work—the idea of integrating many levels of the central dogma (DNA, RNA, and protein) to allow characterization of the heterogeneity in patients to facilitate effective treatment (91).

Peer review: Externally peer-reviewed.

Author Contributions: Concept - B.K.M., M.L.W.; Design - B.K.M., M.E.E., M.H., M.L.W.; Supervision - B.K.M., M.E.E., M.H., M.L.W; Analysis and/or Interpretation - B.K.M., M.E.E., M.H., M.L.W; Literature Search - B.K.M., M.E.E., M.H., M.L.W.; Writing Manuscript - B.K.M., M.E.E., M.H., M.L.W.; Critical Review - B.K.M., M.E.E., M.H., M.L.W.

Conflict of Interest: B.K.M. was supported by T32 AI007363. M.E.E. was supported by T32 AI007363 and Burroughs Wellcome Big Data in the Life Sci-
ence Training Program (PUP 1014106). M.L.W. was supported by grants from the Scleroderma Research Foundation, the Scleroderma Foundation, and from the NIH (PS0AR060780, P44AR072170, P44AR073067, P20GM130454). M.H. was supported by an NIH NIAMS grant (R01AR073270).

Financial Disclosure: The authors declared that this study has received no financial support from industry.

References

1. Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Primers 2015; 1: 15002. [Crossref]

2. Peoples C, Medsger TA, Jr, Lucas M, Rosario BL, Feghali-Bostwick CA. Gender differences in systemic sclerosis: Relationship to clinical features, serologic status and outcomes. J Scleroderma Relat Disord 2016; 1: 177-240. [Crossref]

3. Barnes J, Mayes MD. Epidemiology of systemic sclerosis. Incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol 2012; 24: 165-70. [Crossref]

4. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 2008; 58: 15-25. [Crossref]

5. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: An American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum 2013; 65: 2537-47. [Crossref]

6. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One 2008; 3: e2696. [Crossref]

7. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafayits R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol 2012; 132: 1363-73. [Crossref]

8. Hinrichs C, Pendergrass SA, Sargent JL, Johnson ME, Mahoney JM, Wood TA, Mahoney JM, Kalyanpur A, Wood ML, et al. Expression profile of systemic sclerosis: Understanding pathogenesis and guiding therapy. Arthritis Rheumatol 2019; 71: 392-404. [Crossref]

9. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Plioi PA, et al. Systemic sclerosis: Incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol 2012; 24: 165-70. [Crossref]

10. Allerone Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Primers 2015; 1: 15002. [Crossref]

11. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res 2015; 5: 2929-43. [Crossref]

12. Parkar JS, Mullins M, Cheang M, Leung S, Voduc D, Vickers T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009; 27: 1160-7. [Crossref]

13. Hannoun MB, Breiden-Maskey CL, Raphael J, Brackstone M. A review of gene expression profiling in early stage ER+/HER2- breast cancer with a focus on the PAM50 Risk of Recurrence Assay. EMJ Oncol 2019; 7: 2-13. [Crossref]

14. Johnson ME, Mahoney JM, Taroni J, Sargent JL, Marmarelis E, Wu M-R, et al. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLOS ONE 2015; 10: e0114017. [Crossref]

15. Hinrichs C, Huang C-C, Wood TA, Mahoney JM, Martyanov V, Bhattacharya S, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate monophosphate treatment in systemic sclerosis. J Invest Dermatol 2013; 133: 1979-89. [Crossref]

16. Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafayits R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol 2012; 132: 1363-73. [Crossref]

17. Taroni J, Martyanov V, Huang C-C, Mahoney JM, Hirano I, Shetun B, et al. Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. Arthritis Res Ther 2015; 17: 194. [Crossref]

18. Khanna D, Spino C, Johnson S, Chung L, Whitfield ML, Denton CP, et al. Abatacept in early diffuse cutaneous systemic sclerosis - results of a phase 2 investigator-initiated, multicenter, double-blind randomized placebo-controlled trial. Arthritis Rheumatol 2020; 72: 125-36. [Crossref]

19. Chakravartty EF, Martyanov V, Fiorentino D, Wood TA, Haddon DJ, Jarrell JA, et al. Gene expression changes reflect clinical response in a placebo-controlled randomized trial of abatacept in patients with diffuse cutaneous systemic sclerosis. Arthritis Res Ther 2015; 17: 159. [Crossref]

20. Whitfield ML, Finlay DR, Murray JJ, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell-type-specific gene expression patterns in systemic sclerosis. Proc Natl Acad Sci U S A 2003; 100: 12319-24. [Crossref]

21. Zhou X, Tan FK, Xiong M, Arnett FC, Feghali-Bostwick CA. Monozygotic twins clinically discordant for scleroderma show concordance for fibroblast gene expression profiles. Arthritis Rheum 2005; 52: 3205-14. [Crossref]

22. Derrett-Smith EC, Martyanov V, Chighizola CM, Moinzadeh P, Campochiaro C, Khan K, et al. Limited cutaneous systemic sclerosis skin demonstrates distinct molecular subsets separated by a cardiovascular development gene expression signature. Arthritis Res Ther 2017; 19: 156. [Crossref]

23. Taroni J, Martyanov V, Huang CC, Mahoney JM, Hirano I, Shetun B, et al. Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. Arthritis Res Ther 2015; 17: 194. [Crossref]

24. Feghali-Bostwick CA. Monozygotic twins clinically discordant for scleroderma show concordance for fibroblast gene expression profiles. Arthritis Rheumatol 2015; 67: 3016-26. [Crossref]

25. Sargent JL, Li Z, Alprantis AO, Greenblatt M, Lemaire R, Wu M-H, et al. Identification of optimal mouse models of systemic sclerosis by interspecies comparative genomics. Arthritis Rheumatol 2016; 68: 2003-15. [Crossref]

26. Long KB, Li Z, Burgwin CM, Choe SG, Martyanov V, Sassi-Gaha S, et al. The Tsk2/- mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the Col3a1 gene. J Invest Dermatol 2015; 135: 718-27. [Crossref]

27. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Plioi PA, et al. Systemic sclerosis: Novel insights and therapeutic implications. Curr Rheumatol Rep 2019; 21: 31. [Crossref]

28. Apostolidis SA, Stifano G, Tabibi T, Rice LM, Morse CM, Kaaleheb B, et al. Single cell RNA sequencing identifies HSPG2 and APLN as markers of endothelial cell injury in systemic sclerosis skin. Front Immunol 2018; 9: 2191. [Crossref]

29. Taroni J, Martyanov V, Mahoney JM, Whitfield ML. A functional genomic meta-analysis of clinical trials in systemic sclerosis: Toward precision medicine and combination therapy. J Invest Dermatol 2017; 137: 1033-41. [Crossref]

30. Martyanov V, Whitfield ML. Molecular stratification and precision medicine in systemic sclerosis from genomic and proteomic data. Curr Opin Rheumatol 2016; 28: 83-8. [Crossref]

31. Taroni J, Martyanov V, Mahoney JM, Whitfield ML. A functional genomic meta-analysis of clinical trials in systemic sclerosis: Toward precision medicine and combination therapy. J Invest Dermatol 2017; 137: 1033-41. [Crossref]

32. Chung L, Fiorentino DF, Benbarak MJ, Adler CM, Kahlakeh B, et al. Single cell RNA sequencing identifies HSPG2 and APLN as markers of endothelial cell injury in systemic sclerosis skin. Front Immunol 2018; 9: 2191. [Crossref]

33. Petty E, Feltkamp TR, Callen DF, Chojnowski M, Khanna D, et al. The Tsk2/- mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the Col3a1 gene. J Invest Dermatol 2015; 135: 718-27. [Crossref]

34. Feghali-Bostwick CA. Gender differences in systemic sclerosis. Nat Rev Rheumatol 2012; 8: 118-27. [Crossref]

35. Chung L, Fiorentino DF, Benbarak MJ, Adler CM, Kahlakeh B, et al. Single cell RNA sequencing identifies HSPG2 and APLN as markers of endothelial cell injury in systemic sclerosis skin. Front Immunol 2018; 9: 2191. [Crossref]
37. Gordon JK, Martyanov V, Magro C, Wildman HF, Wood TA, Huang W-T, et al. Nilotinib (Tasigna) in the treatment of early diffuse systemic sclerosis: An open-label, pilot clinical trial. Arthritis Res Ther 2015; 17: 213. [Crossref]

38. Franks JM, Martyanov V, Cai G, Wang Y, Li Z, Wood TA, et al. A machine learning classifier for assigning individual patients with systemic sclerosis to intrinsic molecular subsets. Arthritis Rheumatol 2019; 71: 1701-10. [Crossref]

39. Leijten EFA, Radstake TRDJ, Reedquist KA. Editorial: Lessons learned from a "failed" clinical trial. Arthritis Rheumatol 2018; 70: 1364-5. [Crossref]

40. Feghali-Bostwick C, Medsger TA, Jr., Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. Arthritis Rheum 2003; 48: 1956-63. [Crossref]

41. Radstake TR, Gorlova O, Rueda B, Martin J-E, Alizadeh BZ, Palomo-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010; 42: 426-9. [Crossref]

42. Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R. Analysis of transforming growth factor beta1 gene polymorphisms in patients with systemic sclerosis. Ann Rheum Dis 2002; 61: 678-81. [Crossref]

43. Ohtsuka T, Yamakage A, Yamazaki S. The poly-rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. Arthritis Rheum 2009; 60: 225-33. [Crossref]

44. Gourh P, Agarwal SK, Divecha D, Assassi S, Paz G, Arora-Singh RK, et al. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis. Ann Rheum Dis 2009; 68: 584-90. [Crossref]

45. Zhou X, Tan FK, Reveille JD, et al. The +670G>A polymorphism in the FAS gene promoter region influences the susceptibility to systemic sclerosis. Am Rheum 2009; 68: 700-5. [Crossref]

46. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. J Autoimmun 2010; 34: 155-62. [Crossref]

47. Wipff J, Dieude P, Guedj M, Ruz B, Riemekasten G, Cracowski JL, et al. Association of a KCNA5 gene polymorphism with systemic sclerosis-associated pulmonary arterial hypertension in the European Caucasian population. Arthritis Rheum 2010; 62: 3093-100. [Crossref]

48. Dieude P, Bouaziz M, Guedj M, Riemekasten G, Airé P, Müller 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: 3977-87. [Crossref]

49. Carmona FD, Gutala R, Simeon CP, Carreira P, Ortego-Centeno N, et al. BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. Ann Rheum Dis 2010; 69: 700-5. [Crossref]

50. Zhou X, Tan FK, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]

51. Gourh P, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]

52. Liakouli V, Manetti M, Pacini A, Tolusso B, Fatini M, Ahn C, et al. Association of novel poly-morphisms with susceptibility to and disease prevalence. Arthritis Rheum 2003; 48: 186-92. [Crossref]

53. Gourh P, Agarwal SK, Divecha D, Assassi S, Paz G, Arora-Singh RK, et al. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis. Ann Rheum Dis 2009; 68: 584-90. [Crossref]

54. Rueda B, Gourh P, Broen J, Agarwal SK, Simeon C, Ortego-Centeno N, et al. BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. Ann Rheum Dis 2010; 69: 700-5. [Crossref]

55. Zhou X, Tan FK, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]

56. Rueda B, Gourh P, Broen J, Agarwal SK, Simeon C, Ortego-Centeno N, et al. BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. Ann Rheum Dis 2010; 69: 700-5. [Crossref]

57. Zhou X, Tan FK, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]

58. Zhou X, Tan FK, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]

59. Zhou X, Tan FK, Revelle JD, Wallis D, Milwicz DM, Ahn C, et al. Association of novel poly-morphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002; 46: 2990-9. [Crossref]
75. Wang J, Yi L, Guo X, He D, Li H, Guo G, et al. Lack of association of the CD247 SNP rs2056626 with systemic sclerosis in Han Chinese. Open Rheumatol J 2014; 8: 43-5. [Crossref]

76. Gourh P, Remmers EF, Boyden SE, Alexander T, Morgan ND, Shah AA, et al. Brief report: Whole-exome sequencing to identify rare variants and gene networks that increase susceptibility to scleroderma in African Americans. Arthritis Rheumatol 2018; 70: 1654-60. [Crossref]

77. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol 2015; 11: e1004005. [Crossref]

78. Yang Y, Fan PS, Kahaleh B. Association between enhanced type I collagen expression and epigenetic repression of the FLI1 gene in scleroderma fibroblasts. Arthritis Rheum 2006; 54: 2271-9. [Crossref]

79. Huber LC, Distler JH, Moritz F, Hemmatazad H, Hauser T, Michel BA, et al. Trichostatin A prevents the accumulation of extracellular matrix in a mouse model of bleomycin-induced skin fibrosis. Arthritis Rheum 2007; 56: 2755-64. [Crossref]

80. Lei W, Luo Y, Lei W, Luo Y, Yan K, Zhao S, et al. Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. Scand J Rheumatol 2009; 38: 369-74. [Crossref]

81. Jiang H, Xiao R, Lian X, Kanekura T, Luo Y, Yin Y, et al. Demethylation of TNFSF7 contributes to CD70 overexpression in CD4+ T cells from patients with systemic sclerosis. Clin Immunol 2012; 143: 39-44. [Crossref]

82. Wang Y, Shu Y, Xiao Y, Wang Q, Kanekura T, Li Y, et al. Hypomethylation and overexpression of ITGAL (CD11a) in CD4(+) T cells in systemic sclerosis. Clin Epigenetics 2014; 6: 25. [Crossref]

83. Altorok N, Tsou PS, Coit P, Khanna D, Sawalha AH. Genome-wide DNA methylation analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals common and subset-specific DNA methylation aberrancies. Ann Rheum Dis 2015; 74: 1612-20. [Crossref]

84. Chaudhary P, Chen X, Assassi S, Gorlova O, Dregger H, Harper BE, et al. Cigarette smoking is not a risk factor for systemic sclerosis. Arthritis Rheum 2011; 63: 1098-102. [Crossref]

85. Zhang YJ, Zhang L, Huang XL, Duan Y, Yang LJ, Wang J. Association between cigarette smoking and impaired clinical symptoms in systemic sclerosis: A review. Cell Immunol 2017; 318: 1-7. [Crossref]

86. Dospinescu P, Jones GT, Basu N. Environmental risk factors in systemic sclerosis. Curr Opin Rheumatol 2013; 25: 179-83. [Crossref]

87. Farina A, Cirone M, York M, Lenina S, Padilla C, Mclaughlin S, et al. Epstein-barr virus infection induces aberrant TLR activation pathway and fibroblast-myofibroblast conversion in scleroderma. J Invest Dermatol 2014; 134: 954-64. [Crossref]

88. Bellocchi C, Volkman ER. Update on the gastrointestinal microbiome in systemic sclerosis. Curr Rheumatol Rep 2018; 20: 49. [Crossref]

89. Arron ST, Dimon MT, Li Z, Johnson ME, Wood TA, Feeley L, et al. High rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis. J Invest Dermatol 2014; 134: 2138-45. [Crossref]

90. Johnson ME, Franks JM, Cai G, Mehta BK, Wood TA, Archambault K, et al. Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression in systemic sclerosis skin. Arthritis Res Ther 2019; 21: 49. [Crossref]

91. Hinchcliff M, Mahoney JM. Towards a new classification of systemic sclerosis. Nat Rev Rheumatol 2019; 15: 456-7. [Crossref]