REVIEW

OPEN

Genetics of systemic sclerosis: recent advances

Paula S. Ramos[a,b], *Richard M. Silver*[a], *and Carol A. Feghali-Bostwick*[a]

Purpose of review
Large-scale and follow-up genetic association studies in systemic sclerosis (SSc) have implicated over 40 regions in disease risk, 15 of which with robust associations. Nevertheless, the causal variants and the functional mechanisms underlying the genetic associations remain elusive, and the reasons for the higher disease burden in African Americans unknown. Incorporating tools from diverse fields is beginning to unveil the role of genetic diversity and regulatory variation in SSc susceptibility. This review will summarize recent advances in SSc genetics, including autoimmune disease overlap, evidence of natural selection, and current progress towards the dissection of the functional role of associated risk variants.

Recent findings
In the past year, multiple large-scale studies reported novel strong and suggestive SSc associations. These results, coupled with the regions shared with other autoimmune diseases, emphasize the role of dysregulation of immune pathways as a key causative factor in SSc pathogenesis. Strong evidence implicates natural selection as a mechanism contributing to the maintenance of some of these SSc alleles in the population. Studies integrating genomic, transcriptomic, and epigenomic datasets in specific cell types to identify causal autoimmune disease variants are emerging.

Summary
The identification and comprehensive understanding of the factors and mechanisms contributing to SSc will contribute to improved diagnosis and disease management.

Keywords
functional variation, genetic association, population genetics

INTRODUCTION

Systemic sclerosis (SSc, scleroderma) is a multisystem disease characterized by cutaneous and visceral fibrosis, immune dysregulation, and vasculopathy. SSc is a fibrosing disease preceded and accompanied by autoimmunity and vasculopathy [1]. SSc is very heterogeneous, with patients being commonly classified into two main subsets on the basis of the pattern of skin involvement: limited cutaneous SSc (lcSSc), which is typically associated with anticentromere antibodies (ACA), pulmonary hypertension, but generally an overall better prognosis, and diffuse cutaneous SSc (dcSSc), typically associated with antitopoisomerase I (Scl-70) antibodies (ATA), anti-RNA polymerase antibodies, propensity for severe pulmonary fibrosis, earlier age of onset, and a poorer prognosis [2,3].

There is great variation in reported incidence and prevalence estimates, with the latter varying between 31 and 659 cases per million [4]. Like other autoimmune diseases, SSc is associated with a strong sex bias of between four and nine affected women for one man [5–7]. African Americans are more likely to develop SSc than Caucasians [8,9], and experience greater morbidity, reduced survival, earlier age of onset, and worse pulmonary disease [8–13]. Nearly 70% of SSc patients show pulmonary involvement, which is the main cause of mortality in SSc [14]. The cause of SSc and the reasons for the ethnic disparities remain elusive.

Despite the progress in the identification of genetic risk factors for SSc, little progress has been made in the identification of the specific disease-predisposing functional variations. Since most functional variants lie in noncoding regions of the

[a]Division of Rheumatology and Immunology, Department of Medicine and [b]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

Correspondence to Paula S. Ramos, PhD, Division of Rheumatology and Immunology, Department of Medicine, Medical University of South Carolina, 96 Jonathan Lucas Street, Suite 816, MSC 637, Charleston, SC 29425, USA. Tel: +1 843 792 7081; fax: +1 843 792 7121; e-mail: ramosp@musc.edu

Curr Opin Rheumatol 2015, 27:521–529

DOI:10.1097/BOR.0000000000000214

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

1040-8711 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

KEY POINTS

- Despite its limited genetic burden, multiple genetic risk loci for SSc continue to be identified, underscoring the role of immune system activation as an early and key event in SSc pathogenesis.
- The majority of the SSc-associated loci are shared with at least one other autoimmune disease, mostly with rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, inflammatory bowel disease, and primary biliary cirrhosis.
- Several SSc risk regions show compelling evidence for natural selection, supporting the hypothesis that this population phenomenon has contributed to genetic variation that influences disease risk and ethnic and geographic health disparities.
- Despite the disproportionate burden on African Americans, the dramatic under-representation of African Americans in research continues to exacerbate the health disparities gap.
- The integration of genetic association with population genetic and functional annotation tools will help elucidate the role of recent human adaptation in ethnic disparities and the functional role of associated risk variants in relevant cell types, supporting future gene editing studies aimed at experimentally validating the disease mechanisms.

genome and the majority of complex disease-associated variants are concentrated in regulatory DNA [15], functional studies of associated loci is an area of current focus. In addition, despite the disproportionate burden on African Americans, virtually all studies have been conducted in populations of European ancestry. The dramatic under-representation of African Americans in research continues to exacerbate the health disparities gap and is a major priority. A collaborative large-scale effort to identify common and low-frequency SSc variants in African Americans is underway, which, coupled with the increasing understanding of the regulatory genome, will undoubtedly contribute to the elucidation of the functional mechanisms that underlie SSc cause.

GENETIC RISK

Although multiple lines of indirect evidence support a modest genetic risk in SSc, including a sibling risk ratio (λ_S) of 15–19 [16–19], a low disease concordance in monozygotic twins (4.2%) [20], and genetic heritability ($h = 0.008$) [20], the high concordance of antinuclear antibodies (ANAs) in monozygotic twins (90 and 71% for titers $\geq 1:40$ and $\geq 1:80$, respectively) [20] and the increased relative risk of autoimmune disease in first-degree relatives [21] support a higher genetic contribution to the development of autoimmunity in SSc. The modest genetic burden suggests a substantial role for epigenetic or environmental factors in SSc susceptibility.

Recently, more direct evidence for the role of genetic variation in the pathogenesis SSc has emerged. Until 2010, only a handful of candidate gene polymorphisms had been convincingly implicated is SSc risk. With the advent and application of high-throughput technologies and increased collaborations, four genome-wide association studies (GWAS) [22–25], two large-scale studies [26▪,27▪] using a high-density genotyping array of immune loci (Immunochip), and multiple GWAS and Immunochip follow-up studies [28–32] have been published since 2010. Currently, 45 genetic regions are implicated in SSc risk, 15 of which have compelling genetic associations at the genome-wide significance level ($P < 5 \times 10^{-8}$) and/or multiple suggestive ($5 \times 10^{-8} < P < 5 \times 10^{-4}$) independent replications. These regions are shown in Table 1, which, in order to avoid likely spurious associations, includes all genetic associations with minimum P value no larger than 5×10^{-4} tested in a total sample size of at least 1000 individuals. In spite of the complex genetic architecture, these discoveries demonstrate that a broad array of pathways underlies the genetic heterogeneity of SSc, including innate immunity pathways (Toll-like receptor, interferon, DNA clearance during apoptosis), adaptive immunity pathways [interleukin (IL)-12], and fibrosis pathways.

The human leukocyte antigen (HLA) class II region remains the most significant genetic association with SSc, similar to all other autoimmune diseases, and has been extensively reviewed [70]. Non-HLA loci have also been reviewed [71,72]. In the past year, two Immunochip [26▪,27▪], an Immunochip and GWAS follow-up [31,32], and a candidate gene study [73] reported novel strong and suggestive associations with SSc. The findings from these studies are discussed below, some of which have already been included in recent reviews [74,75]. The majority of these regions overlap with those implicated in other autoimmune diseases, underscoring the role of immune system dysregulation as a primary event in SSc pathogenesis.

Epigenetic studies [76,77] have been recently reviewed elsewhere and are not discussed here.

NEW STUDIES

Beyond the initial wave of GWAS studies, gene identification studies have since aimed to capture as much genomic variation as possible in each region to maximize the odds of identifying the

Table 1. Genetic regions associated with systemic sclerosis, including evidence for sharing with autoimmune diseases, for natural selection, and implicated agent of selection

Gene region	Band location	Approach	Phenotype	References	AD overlap	Natural selection	Selective pressure
KIAA0319L	1p34.3	GWA	lcSSc	[24]			
IL12RB2*	1p31.3	GFU		[28]	BeD, PBC		
IL23R	1p31.3	CG		[33,34]	AS, BeD, CD, IBD, PS, UC	Y	Protozoa
VCAM1	1p21.2	IC		[27▪]	MS		
PTPN22	1p13.2	CG		[35–37]	CD, MG, RA, T1D, VT	Y	Protozoa
CD247*	1q24.2	CG, GWA		[22,23,25,38]	CelD, RA		
TNFSF4*	1q25.1	CG, GWA	ACA	[24,39–41]	MS, RA, SLE	Y	
NMNAT2	1q25.3	GWA	ACA	[24]			
RHOB	2p24.1	GWA		[22]			
STAT4*	2q32.3	CG, GWA		[22–25,27▪,42–45]	BeD, IBD,PBC, RA, SLE, SS, CelD + RA		
PPARG	3p25	GFU		[32]			
PLCL2	3p24.3	CG		[73]	MS, PBC, RA		
DNASE1L3-PXK*	3p14.1	IC	ACA	[24,26▪,27▪]	RA, SLE		
IL12A*	3q25.33	IC	lcSSc	[26▪]	BeD, CelD, MS, PBC	Y	
TLR2	4q32	CG	ATA	[46]			
BANK1-NFKB1	4q24	CG, GFU	dcSSc, ATA	[29,73,47,48]	CD, PBC, MS, SLE, UC		
TNIP1*	5q33.1	GFU, GWA		[22,24,30]	IBD, MG, PS, SLE	Y	
HLA region*	6p21.33			[70]	All ADs	Y	Bacterial infection
UHRF1BP1	6p21.31	GWA		[24]	SLE	Y	*M. tuberculosis*
ATG5*	6q21	IC, GWA		[24,26▪]	RA, SLE		
TNFAIP3*	6q23.3	CG, GWA	SSc, dcSSc and ATA	[24,49–51]	CelD, IBD, MS, PS, RA, SLE, SS, UC		
CCR6	6q27	CG	ATA	[52]	RA		
JAZF1	7p15.1	GWA		[24]	CD, RA, T1D		
GRB10	7p12.1	GWA	lcSSc	[23]			
IRF5*	7q32.1	CG, GWA		[22–25,27▪,53–55]	PBC, RA, SLE, UC		
BLK*	8p23.1	CG, GWA		[24,56–58]	KA, RA, SLE	Y	
PSD3	8p21.3	GFU		[29]			
IL2RA	10p15.1	CG	ACA	[59]	AA, CD, IBD, MS, RA, T1D		
FAS	10q24.1	CG		[60,61]			
IRF7	11p15.5	CG, GWA	ACA	[24,62]	SLE		

(*Continued*)

1040-8711 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Table 1 (*Continued*)

Gene region	Band location	Approach	Phenotype	References	AD overlap	Natural selection	Selective pressure
EHF	11p13	GWA	lcSSc	[24]	SLE		
TREH-DDX6	11q23.3	IC		[26▪]	SLE, CelD + RA		
KCNA5	12p13	CG	SScPAH	[63]			
SOX5	12p12.1	GWA	ACA	[23]			
RPL41-ESYT1	12q13.2	GWA	dcSSc	[23]	T1D		
PLD4	14q32.33	CG		[50]	RA		
CSK*	15q24.2	GFU, GWA		[24,29]			
IRF8*	16q24.1	GWA	lcSSc	[23,24,73]	IBD, MS, PBC, RA, UC	Y	
NLRP1	17p13.2	CG		[64]		Y	
IKZF3	17q12	GWA	ACA	[24]	IBD, PBC		
TBX21	17q21.32	CG		[64]	AS, MS		
CD226	18q22.2	CG		[65]	IBD, RA, T1D		
IL12RB1*	19p13.1	IC FU		[31]	MS		
MIF	22q11.23	CG		[66,67]			
IRAK1	Xq28	CG		[68,69]	RA		

Disease abbreviations: AA, alopecia areata; AD, autoimmune disease; AS, ankylosing spondylitis; BeD, Behçet's disease; CelD, celiac disease; CD, Crohn's disease; IBD, inflammatory bowel disease; KA, Kawasaki disease; MG, myasthenia gravis; MS, multiple sclerosis; PBC, primary biliary cirrhosis; PS, psoriasis; RA, rheumatoid arthritis; HLA: human leukocyte antigen; SLE, systemic lupus erythematosus; SS, Sjögren's syndrome; T1D, type 1 diabetes; UC, ulcerative colitis; VT, vitiligo. No overlap was observed for Addison's disease, autoimmune thyroid disease, Graves' disease, juvenile idiopathic arthritis, primary sclerosing cholangitis, psoriatic arthritis, and sarcoidosis. Genetic regions with strong evidence of selection attributed to host–pathogen coevolution were compiled from a recent review [86▪].

*Established SSc loci (genome-wide significance or two independent replications with $P < 5 \times 10^{-4}$) are denoted by an asterisk. Approaches used include genome-wide association (GWA), candidate gene (CG), Immunochip (IC), GWAS follow-up (GFU), and IC follow-up (IC FU) studies. SSc phenotypes with stronger association than SSc are shown (lcSSC: limited cutaneous SSc; dcSSC: diffuse cutaneous SSc; ATA: antitopoisomerase I antibody; ACA: anticentromere antibody; SScPAH: SSc-associated pulmonary arterial hypertension). Autoimmune disease (AD) sharing based on genetic regions reported in the National Human Genome Research Institute's Catalog of Published Genome-Wide Association Studies (http://www.genome.gov/gwastudies) accessed on 27 March 2015 [80].

potential causal variants. Approaches include fine mapping, resequencing, and imputation of variants not directly genotyped. Since the first GWAS and, more recently, Immunochip studies, a plethora of manuscripts following up individual candidate genes in independent cohorts have been published, sometimes with statistical significance below suggestive evidence for association. Hence, caution must be used when interpreting association results.

Early in 2014, the results of the first Immunochip study [26▪] were published, unveiling strong SSc associations in the *IL12A*, *ATG5*, *DNASE1L3-PXK*, and *TREH-DDX6* regions. At the same time, a GWAS follow-up study [32] reported a suggestive association with *PPARG*. Later in the year, a second Immunochip study confirmed the association in the *DNASE1L3-PXK* region and reported a novel suggestive association with *VCAM1* [27▪]. More recently, an Immunochip follow-up study [31] reported a strong association with *IL12RB1*, and a candidate gene study [73] reported suggestive evidence of association between *PLCL2* and SSc.

As with the vast majority of all reported associations, the causal variants remain unknown. Only the signal in the *DNASE1L3-PXK* region is most likely to be driven by a missense variant in DNASE1L3 [26▪,27▪]. Consistently with the association of ACA antibody status with lcSSc, this association is stronger in ACA-positive as well as lcSSc cases. Given the role of DNASE1L3 in DNA degradation in apoptosis, loss of function of the protein and consequent defects in DNA clearance might contribute to the production of ACA [26▪,27▪].

Since, by design, the Immunochip array targets immune-related genes, it is not surprising that most of the implicated genes participate in immune processes. IL12A and ATG5 are involved in interferon signaling, and ATG5 has roles in autophagy. IL12RB1 is a component of the IL12 pathway. VCAM1 mediates leukocyte–endothelial cell adhesion and signal transduction. PLCL2 participates in B-cell receptor signaling and B-cell proliferation. Nevertheless, one of these recently reported genes is involved in fibrosis: PPARG is an antifibrotic effector whose expression may influence the uncontrolled progression of fibrosis in SSc [32]. Collectively, these associations corroborate the role for defects in DNA clearance, autophagy, interferon signaling, IL12 signaling, B-cell signaling, cell adhesion, and fibrotic processes in disease susceptibility.

Clearly, although fibroblast activation is the hallmark of SSc, most of the genetic factors associated with SSc lie in immune-related genes [71,75], and the reasons underlying the characteristic excess deposition of extracellular matrix (ECM) proteins remain elusive. The distinct genetic architecture of pulmonary fibrosis in SSc-associated interstitial lung disease compared with idiopathic pulmonary fibrosis is noteworthy, supporting a distinct genetic risk of these two forms of lung fibrosis and suggesting that fibrosis might be the result of different pathogenic mechanisms, with a more immune-driven cause in SSc [78]. In a recent bioinformatic analysis of gene expression data from skin biopsies integrated with SSc-associated genetic variation, Mahoney *et al.* [79▪] corroborated that the genetic risk in SSc was overwhelmingly related to immune abnormalities. Their data further support the hypothesis that immune system activation is an early and key event in SSc pathogenesis likely involving interferon activation and macrophage recruitment, which may influence or drive ECM remodeling and cell proliferation in the skin [79▪].

SHARED LOCI WITH OTHER AUTOIMMUNE DISEASES

The genetic regions currently reported in the National Human Genome Research Institute (NHGRI) GWAS catalog [80] as being associated with an autoimmune disease are included in Table 1. Two-thirds (30/45) of the SSc-associated loci are also associated with at least one other autoimmune disease, 46% (21/45) with at least two diseases, 37% (16/45) with three, 24% (11/45) with four, 17% (8/45) with five, and 11% (5/45) with at least six autoimmune diseases. Since we are only including loci reported in the NHGRI GWAS catalog, which does not include all the regions with established associations with autoimmune diseases, these numbers are conservative. The non-HLA loci that SSc shares with most autoimmune diseases include *TNFAIP3* (eight other autoimmune diseases), *IL23R* (six autoimmune diseases), *STAT4* (six autoimmune diseases), and *IL2RA* (six autoimmune diseases). Interestingly, despite having been identified in GWAS of six different autoimmune diseases, the IL21 region has been reported as modestly associated with SSc [81], and hence did not meet the criteria to be included in Table 1. The diseases with which SSc shares the most pleiotropic effects include rheumatoid arthritis (16 genetic regions), systemic lupus erythematosus (SLE) (13 regions), multiple sclerosis (9 regions), inflammatory bowel disease (8 regions), and primary biliary cirrhosis (PBC) (8 regions). A recent clustering of autoimmune diseases based on the number of shared genetic loci using a more stringent genome-wide significance level showed SSc to be highly correlated with PBC and SLE [82▪▪]. Given the known clinical associations of SSc

1040-8711 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

with PBC [83] or SLE [84], these findings are not surprising.

It is thought that comparison with PBC, an autoimmune liver fibrotic disorder, might help uncover the genetic risks underlying excessive fibrosis. A total of 25 loci from four GWAS are currently reported in the NHGRI GWAS catalog as being associated with PBC. As shown in Table 1, all the genetic regions shared between SSc and PBC harbor genes whose primary functions are immune-related (*IRF5, IRF8, IKZF3, STAT4, IL12A, IL12RB2, BANK1-NFKB, PLCL2*), providing further support for the main role of dysregulated immune pathways in these fibrotic disorders.

POPULATION GENETIC TOOLS

The genetic basis of disease is influenced by individual and population variation. The reasons for the geographic and ethnic disparities in the prevalence of SSc and the relative high frequency of SSc risk alleles in the population are not fully understood. Population genetic factors such as natural selection alter allele frequencies over generations and may help explain the persistence of such common risk variants in the population and the differential risk of SSc. For example, in an analysis of correlations between genetic risk of multiple diseases and worldwide migration trajectories, the authors report that variants associated with, among other, SSc, PBC, and SLE, have undergone genetic risk differentiation associated with migration [85].

As we have recently reviewed [86▪], immune responses can be particularly sensitive to environmental change, and immune function genes and pathways are consistently reported in tests for natural selection. Since infectious organisms are strong agents of natural selection, it is thought that the adaptation to pathogen pressure through functional variation in immune-related genes conferred a specific selective advantage for host survival, including protection from pathogens and tolerance to microbiota [87]. Interestingly, since fibrosis has a beneficial role in limiting pathogen invasion and promoting wound healing in response to injury, it has been hypothesized that fibrosis is also an evolutionarily conserved adaptive process [88]. Given the key role of immune and inflammatory processes in the cause of SSc, it is thus important to understand how population genetic factors such as natural selection have contributed to genetic variation that influences disease risk and ethnic and geographic health disparities in individuals and populations.

We have recently summarized the evidence for autoimmune disease-associated loci under selection and the candidate selective pressures [86▪]. As shown in Table 1, 10 SSc risk regions show strong evidence for selection, including the *HLA* region, *PTPN22,* IL23R, *TNFSF4, IL12A, TNIP1*, *UHRF1BP1, BLK, IRF8*, and *NLRP1* [86▪]. Evidence of selection that has been attributed to host–pathogen coevolution is also shown in Table 1. In four of the SSc-associated regions with evidence for selection, known pathogens have been implicated as the selective pressure. These include, for example, variation in the *HLA* as a protective factor against bacterial infection, and resistance to protozoa infection as the selective pressure for *PTPN22*. It is noteworthy that natural selection helps explain the pleiotropy and complex genetic patterns observed at loci like PTPN22, where the same nonsynonymous variant increases the risk of some, but is protective against other autoimmune diseases [86▪].

Regardless of the population phenomenon and the reasons for the emergence of both common and rare SSc-causing alleles, genetic variation is the basis for clinically relevant traits at both the individual and population levels, and thus incorporating population genetics to understand human genetic diversity will lead to a better understanding of the causes of health disparities, identification of functional variants and discovery of cellular mechanisms, and contribute to the development of new therapies.

EMERGING STUDIES

As mentioned above, large-scale studies have greatly enhanced our understanding of the complex cause of SSc, but a full understanding of its genetic etiology is unclear. Since these studies were designed to detect common variants, the role of rare variation remains unknown. Also, these studies were conducted in populations of European ancestry; hence variation in other ethnic groups remains elusive. The National Human Genome Institute [National Institute of Health (NIH)/NHGRI] is leading a large collaborative project entitled 'Genome Research in African American Scleroderma Patients' (GRASP Study) with the goal of discovering common and low-frequency variants associated with SSc susceptibility in African Americans.

The role of nonadditive inter-relationships between loci and the effect of environmental exposures in modulating these allelic effects remain unexplored. The role for epigenetic effects is now beginning to be pursued. Research on the role of the microbiome in SSc is still lacking. Variation in microbiome structure has been associated with other autoimmune diseases, and there is an increasing appreciation for the impact of the host genotype

on the microbiome, and the impact of the microbiome on host. Finally, experimental validation of these genetic associations is lacking. Although functional causal alleles can be protein-altering, the majority of genetic changes associated with complex diseases are concentrated in regulatory DNA [15]. Despite this central importance of noncoding variants, the lack of experimentally validated regulatory mechanisms is hampering the understanding of the regulatory mechanisms underlying observed genotype–phenotype associations. Due to the advancements of sequencing technologies in the past few years, functional genomics, epigenomics, and regulatory genomics datasets are beginning to allow the dissection of the most likely regulatory variants, so that these can be experimentally validated.

For example, Farh *et al.* [82▪▪] integrated genetic and epigenetic data, including transcription and cis-regulatory element annotations in primary immune cells, to identify causal autoimmune disease variants, including SSc. They found that approximately 90% of causal variants were noncoding, 60% mapped to immune-cell enhancer elements, with causal variants often occurring near binding sites for master regulators of immune differentiation and stimulus-dependent gene activation. The authors report that the enhancers and promoters mapping to SSc-associated variants did not reveal strong enrichment of specific cell subsets, but a modest enrichment of T-cell types, in a pattern more similar to vitiligo or primary sclerosing cholangitis [82▪▪]. In contrast, both SLE and PBC variants preferentially mapped to enhancers and promoters active in B cells [82▪▪]. Since this study only included a limited number of SSc-associated regions that met genome-wide significance level in the NHGRI GWAS catalog (13 regions) and did not include any nonimmune cell or tissue that might be relevant in SSc (e.g. skin, lung, vasculature), the results might be a consequence of the limited number of SSc-associated loci, and/or cell types analyzed. The authors concluded that the noncoding variants might cause subtle key differences in transcription or epigenetics that influence autoimmunity.

The identification of the specific cell types most relevant to disease biology is an area of major interest. This year, the NIH Roadmap Epigenomics Consortium published an integrative analysis of 111 reference human epigenomes profiled for histone modification patterns, DNA accessibility, DNA methylation, and RNA expression [89▪▪]. SSc was not included; however, the authors reported that genetic variants associated with SLE were the most enriched in H3 lysine 4 monomethylation (H3K4me1) peaks, a marker of enhancer regions, in memory T-helper and B cells, whereas genetic variants associated with PBC were the most enriched in this enhancer mark in B cells and lymphocytes [89▪▪]. Looking at the relevant SSc tissues, fibroblasts were the most enriched with enhancer marks for height-associated variants, and vascular tissue was enriched with aortic root size variants. These results illustrate that tissue-specific epigenomic annotations can help to identify biologically relevant cell types and to interpret genetic changes associated with disease.

CONCLUSION

In spite of the limited genetic burden, multiple genetic risk loci for SSc continue to be identified, underscoring the role of immune pathways in its genetic risk. Nevertheless, the identification of the specific disease-predisposing functional variations and reasons underlying the ethnic disparities remain unknown. African Americans suffer worse health outcomes in part because they are severely underrepresented in research and, consequently, may be omitted from developments in new genetic technologies and clinical advancements. This disengagement in research participation can only serve to exacerbate the health disparities gap as healthcare moves towards precision medicine, and must imperatively be addressed.

The integration of genetic association with population genetic and functional annotation tools will help elucidate the role of recent human adaptation in ethnic disparities and the functional role of associated risk variants in relevant cell types. This knowledge will guide gene editing studies that will uncover the disease mechanisms, which has direct implications for personalized medicine approaches.

Acknowledgements

None.

Financial support and sponsorship

This study was supported by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (NIH) under Award Numbers K01 AR067280, K24 AR060297, P30 AR061271, R03 AR065801, R21 AR065089, P60 AR062755, and by the South Carolina Clinical & Translational Research (SCTR) Institute, with an academic home at the Medical University of South Carolina, through NIH Grants nos UL1 RR029882 and UL1 TR000062. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest

There are no conflicts of interest.

1040-8711 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Varga J, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. J Clin Invest 2007; 117:557–567.
2. Steen VD. The many faces of scleroderma. Rheum Dis Clin North Am 2008; 34:1–15; v.
3. LeRoy EC, Black C, Fleischmajer R, *et al.* Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988; 15:202–205.
4. Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol 2012; 24:165–170.
5. Meier FM, Frommer KW, Dinser R, *et al.* Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. Ann Rheum Dis 2012; 71:1355–1360.
6. Ferri C, Valentini G, Cozzi F, *et al.* Systemic sclerosis: demographic, clinical, and serologic features and survival in 1012 Italian patients. Medicine (Baltimore) 2002; 81:139–153.
7. Scussel-Lonzetti L, Joyal F, Raynauld JP, *et al.* Predicting mortality in systemic sclerosis: analysis of a cohort of 309 French Canadian patients with emphasis on features at diagnosis as predictive factors for survival. Medicine (Baltimore) 2002; 81:154–167.
8. Laing TJ, Gillespie BW, Toth MB, *et al.* Racial differences in scleroderma among women in Michigan. Arthritis Rheum 1997; 40:734–742.
9. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, *et al.* Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003; 48:2246–2255.
10. Greidinger EL, Flaherty KT, White B, *et al.* African-American race and antibodies to topoisomerase I are associated with increased severity of scleroderma lung disease. Chest 1998; 114:801–807.
11. Krishnan E, Furst DE. Systemic sclerosis mortality in the United States: 1979–1998. Eur J Epidemiol 2005; 20:855–861.
12. McNearney TA, Reveille JD, Fischbach M, *et al.* Pulmonary involvement in systemic sclerosis: associations with genetic, serologic, sociodemographic, and behavioral factors. Arthritis Rheum 2007; 57:318–326.
13. Nietert PJ, Mitchell HC, Bolster MB, *et al.* Racial variation in clinical and immunological manifestations of systemic sclerosis. J Rheumatol 2006; 33:263–268.
14. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972–2002. Ann Rheum Dis 2007; 66:940–944.
15. Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. Science 2012; 337:1190–1195.
16. Hudson M, Rojas-Villarraga A, Coral-Alvarado P, *et al.* Polyautoimmunity and familial autoimmunity in systemic sclerosis. J Autoimmun 2008; 31:156–159.
17. Englert H, Small-McMahon J, Chambers P, *et al.* Familial risk estimation in systemic sclerosis. Aust N Z J Med 1999; 29:36–41.
18. Arora-Singh RK, Assassi S, del Junco DJ, *et al.* Autoimmune diseases and autoantibodies in the first degree relatives of patients with systemic sclerosis. J Autoimmun 2010; 35:52–57.
19. Arnett FC, Cho M, Chatterjee S, *et al.* Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis Rheum 2001; 44:1359–1362.
20. 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; 48:1956–1963.
21. Frech T, Khanna D, Markewitz B, *et al.* Heritability of vasculopathy, autoimmune disease, and fibrosis in systemic sclerosis: a population-based study. Arthritis Rheum 2010; 62:2109–2116.
22. 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:e1002091.
23. 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:e1002178.
24. 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; 22:4021–4029.
25. 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; 42:426–429.
26. ■ Mayes MD, Bossini-Castillo L, Gorlova O, *et al.* Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet 2014; 94:47–61.

First large-scale study using a high-density genotyping array of immune loci (Immunochip), uncovered multiple strong SSc associations.

27. ■ Zochling J, Newell F, Charlesworth JC, *et al.* An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. Arthritis Res Ther 2014; 16:438.

Second large-scale Immunochip study suggests a missense variant in the *DNASE1L3* gene as causal.

28. 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; 21:926–933.
29. 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; 21:2825–2835.
30. Bossini-Castillo L, Martin JE, Broen J, *et al.* Confirmation of TNIP1 but not RHOB and PSORS1C1 as systemic sclerosis risk factors in a large independent replication study. Ann Rheum Dis 2013; 72:602–607.
31. Lopez-Isac E, Bossini-Castillo L, Guerra SG, *et al.* Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus. Arthritis Rheumatol 2014; 66:3521–3523.
32. Lopez-Isac E, Bossini-Castillo L, Simeon CP, *et al.* A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility. Arthritis Res Ther 2014; 16:R6.
33. Agarwal SK, Gourh P, Shete S, *et al.* Association of interleukin 23 receptor polymorphisms with antitopoisomerase-I positivity and pulmonary hypertension in systemic sclerosis. J Rheumatol 2009; 36:2715–2723.
34. Rueda B, Broen J, Torres O, *et al.* The interleukin 23 receptor gene does not confer risk to systemic sclerosis and is not associated with systemic sclerosis disease phenotype. Ann Rheum Dis 2009; 68:253–256.
35. Gourh P, Tan FK, Assassi S, *et al.* Association of the PTPN22 R620W polymorphism with antitopoisomerase I- and anticentromere antibody-positive systemic sclerosis. Arthritis Rheum 2006; 54:3945–3953.
36. 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; 58:2183–2188.
37. Diaz-Gallo LM, Gourh P, Broen J, *et al.* Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Ann Rheum Dis 2011; 70:454–462.
38. 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; 70:1695–1696.
39. Gourh P, Arnett FC, Tan FK, *et al.* Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis. Ann Rheum Dis 2010; 69:550–555.
40. 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; 70:638–641.
41. 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; 39:997–1003.
42. Tsuchiya N, Kawasaki A, Hasegawa M, *et al.* Association of STAT4 polymorphism with systemic sclerosis in a Japanese population. Ann Rheum Dis 2009; 68:1375–1376.
43. Rueda B, Broen J, Simeon C, *et al.* The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet 2009; 18:2071–2077.
44. 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; 60:3794–3806.
45. 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; 60:2472–2479.
46. Broen JC, Bossini-Castillo L, van Bon L, *et al.* A rare polymorphism in the gene for Toll-like receptor 2 is associated with systemic sclerosis phenotype and increases the production of inflammatory mediators. Arthritis Rheum 2012; 64:264–271.
47. 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; 60:3447–3454.
48. 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; 69:700–705.
49. 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; 69:1958–1964.
50. Terao C, Ohmura K, Kawaguchi Y, *et al.* PLD4 as a novel susceptibility gene for systemic sclerosis in a Japanese population. Arthritis Rheum 2013; 65:472–480.
51. 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; 64:2746–2752.
52. Koumakis E, Bouaziz M, Dieude P, *et al.* A regulatory variant in CCR6 is associated with susceptibility to antitopoisomerase-positive systemic sclerosis. Arthritis Rheum 2013; 65:3202–3208.

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; 60:1845–1850.

54. 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; 60:225–233.

55. 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; 37:987–992.

56. 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; 34:155–162.

57. Ito I, Kawaguchi Y, Kawasaki A, *et al.* Association of the FAM167A-BLK region with systemic sclerosis. Arthritis Rheum 2010; 62:890–895.

58. 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; 63:2091–2096.

59. 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; 13:191–196.

60. Liakouli V, Manetti M, Pacini A, *et al.* The -670G>A polymorphism in the FAS gene promoter region influences the susceptibility to systemic sclerosis. Ann Rheum Dis 2009; 68:584–590.

61. Broen J, Gourh P, Rueda B, *et al.* The FAS -670A>G polymorphism influences susceptibility to systemic sclerosis phenotypes. Arthritis Rheum 2009; 60:3815–3820.

62. 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; 71:114–119.

63. Wipff J, Dieude P, Guedj M, *et al.* Association of a KCNA5 gene polymorphism with systemic sclerosis-associated pulmonary arterial hypertension in the European Caucasian population. Arthritis Rheum 2010; 62:3093–3100.

64. 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; 70:668–674.

65. 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; 63:1097–1105.

66. Wu SP, Leng L, Feng Z, *et al.* Macrophage migration inhibitory factor promoter polymorphisms and the clinical expression of scleroderma. Arthritis Rheum 2006; 54:3661–3669.

67. Bossini-Castillo L, Simeon CP, Beretta L, *et al.* Confirmation of association of the macrophage migration inhibitory factor gene with systemic sclerosis in a large European population. Rheumatology (Oxford) 2011; 50:1976–1981.

68. 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; 63:3979–3987.

69. Carmona FD, Cenit MC, Diaz-Gallo LM, *et al.* New insight on the Xq28 association with systemic sclerosis. Ann Rheum Dis 2013; 72:2032–2038.

70. Luo Y, Wang Y, Wang Q, *et al.* Systemic sclerosis: genetics and epigenetics. J Autoimmun 2013; 41:161–167.

71. Martin JE, Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. Hum Genet 2012; 131:1023–1037.

72. Assassi S, Radstake TR, Mayes MD, *et al.* Genetics of scleroderma: implications for personalized medicine? BMC Med 2013; 11:9.

73. Arismendi M, Giraud M, Ruzehaji N, *et al.* Identification of NF-kappaB and PLCL2 as new susceptibility genes and highlights on a potential role of IRF8 through interferon signature modulation in systemic sclerosis. Arthritis Res Ther 2015; 17:71.

74. Korman BD, Criswell LA. Recent advances in the genetics of systemic sclerosis: toward biological and clinical significance. Curr Rheumatol Rep 2015; 17:21.

75. Jin J, Chou C, Lima M, *et al.* Systemic sclerosis is a complex disease associated mainly with immune regulatory and inflammatory genes. Open Rheumatol J 2014; 8:29–42.

76. Altorok N, Almeshal N, Wang Y, *et al.* Epigenetics, the holy grail in the pathogenesis of systemic sclerosis. Rheumatology (Oxford) 2014. [Epub ahead of print]

77. Broen JC, Radstake TR, Rossato M. The role of genetics and epigenetics in the pathogenesis of systemic sclerosis. Nat Rev Rheumatol 2014; 10:671–681.

78. Herzog EL, Mathur A, Tager AM, *et al.* Review: interstitial lung disease associated with systemic sclerosis and idiopathic pulmonary fibrosis: how similar and distinct? Arthritis Rheumatol 2014; 66:1967–1978.

79. ▪ Mahoney JM, Taroni J, Martyanov V, *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.

Integration of genetic association with gene expression data from skin biopsies corroborated that the genetic risk in SSc was overwhelmingly related to immune abnormalities.

80. Hindorff LA, MacArthur J, Morales J, *et al.* A catalog of published genome-wide association studies. www genome gov/gwastudies. [Accessed 27 March 2015]

81. 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; 72:1233–1238.

82. ▪▪ Farh KK, Marson A, Zhu J, *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 2015; 518:337–343.

Comprehensive integration of genetic and epigenetic data, including transcription and cis-regulatory element annotations in primary immune cells, to identify causal autoimmune disease variants, including SSc. This study shows that approximately 90% of causal variants are noncoding, which might cause subtle key differences in transcription or epigenetics that influence autoimmunity.

83. Rigamonti C, Bogdanos DP, Mytilinaiou MG, *et al.* Primary biliary cirrhosis associated with systemic sclerosis: diagnostic and clinical challenges. Int J Rheumatol 2011; 2011:976427.

84. Rojas-Villarraga A, Amaya-Amaya J, Rodriguez-Rodriguez A, *et al.* Introducing polyautoimmunity: secondary autoimmune diseases no longer exist. Autoimmune Dis 2012; 2012:254319.

85. Corona E, Chen R, Sikora M, *et al.* Analysis of the genetic basis of disease in the context of worldwide human relationships and migration. PLoS Genet 2013; 9:e1003447.

86. ▪ Ramos PS, Shedlock AM, Langefeld CL. Genetics of autoimmune diseases: insights from population genetics. J Hum Genet 2015; doi: 10.1038/jhg.2015.94. [Epub ahead of print]

This review summarizes the genetic regions associated with susceptibility to different autoimmune diseases and concomitant evidence for selection, including known agents of selection exerting selective pressure in these regions. It highlights how population genetics phenomena have contributed to genetic variation that influences autoimmune disease risk.

87. Quintana-Murci L, Clark AG. Population genetic tools for dissecting innate immunity in humans. Nat Rev Immunol 2013; 13:280–293.

88. Thannickal VJ, Zhou Y, Gaggar A, *et al.* Fibrosis: ultimate and proximate causes. J Clin Invest 2014; 124:4673–4677.

89. ▪▪ Roadmap Epigenomics C, Kundaje A, Meuleman W, *et al.* Integrative analysis of 111 reference human epigenomes. Nature 2015; 518:317–330.

The NIH Roadmap Epigenomics Consortium integrated histone modification patterns, DNA accessibility, DNA methylation, and RNA expression data to establish global maps of regulatory elements, define regulatory modules of coordinated activity and their likely activators and repressors. This study shows that disease and trait-associated genetic variants are enriched in tissue-specific epigenomic marks, revealing biologically relevant cell types for diverse human traits, and providing a resource for interpreting the molecular basis of human disease.

1040-8711 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.