

Genetic Interactions Affect Lung Function in Patients with Systemic Sclerosis

Anna Tyler,* J. Matthew Mahoney,^{†,*} and Gregory W. Carter*^{‡,1}

*The Jackson Laboratory, 600 Main St. Bar Harbor, ME, [†]Department of Neurological Sciences, Larner College of Medicine, University of Vermont, Burlington, VT, and [‡]Department of Computer Science, University of Vermont, Burlington, VT

ORCID IDs: 0000-0001-8371-2377 (A.T.); 0000-0003-1425-5939 (J.M.M.); 0000-0002-2834-8186 (G.W.C.)

ABSTRACT Scleroderma, or systemic sclerosis (SSc), is an autoimmune disease characterized by progressive fibrosis of the skin and internal organs. The most common cause of death in people with SSc is lung disease, but the pathogenesis of lung disease in SSc is insufficiently understood to devise specific treatment strategies. Developing targeted treatments requires not only the identification of molecular processes involved in SSc-associated lung disease, but also understanding of how these processes interact to drive pathology. One potentially powerful approach is to identify alleles that interact genetically to influence lung outcomes in patients with SSc. Analysis of interactions, rather than individual allele effects, has the potential to delineate molecular interactions that are important in SSc-related lung pathology. However, detecting genetic interactions, or epistasis, in human cohorts is challenging. Large numbers of variants with low minor allele frequencies, paired with heterogeneous disease presentation, reduce power to detect epistasis. Here we present an analysis that increases power to detect epistasis in human genome-wide association studies (GWAS). We tested for genetic interactions influencing lung function and autoantibody status in a cohort of 416 SSc patients. Using Matrix Epistasis to filter SNPs followed by the Combined Analysis of Pleiotropy and Epistasis (CAPE), we identified a network of interacting alleles influencing lung function in patients with SSc. In particular, we identified a three-gene network comprising *WNT5A*, *RBMS3*, and *MSI2*, which in combination influenced multiple pulmonary pathology measures. The associations of these genes with lung outcomes in SSc are novel and high-confidence. Furthermore, gene coexpression analysis suggested that the interactions we identified are tissue-specific, thus differentiating SSc-related pathogenic processes in lung from those in skin.

KEYWORDS

Epistasis
Systemic
Sclerosis
Scleroderma
Wnt Signaling

Systemic sclerosis (SSc), or scleroderma, is a complex autoimmune disease associated with significant morbidity and mortality. It is characterized by widespread fibrosis of skin and internal organs, as well as vasculopathy (Denton and Khanna 2017). The most common cause of death in SSc is lung disease (Steen and Medsger 2007). Roughly a

quarter of people with SSc develop interstitial lung disease, which involves lung fibrosis, vascular hyperreactivity, and inflammation (Solomon *et al.* 2013). Another 7–13% of patients develop pulmonary arterial hypertension, which is characterized by vascular injury and occlusion, vasoconstriction, and dysregulated angiogenesis (Solomon *et al.* 2013). Both conditions lead to reduced lung function and increased risk of death. The pathogenesis of lung disease in SSc is not understood well enough for development of specific treatments, and current treatments rely primarily on non-specific immune suppression (Cappelli *et al.* 2015). There is a need to identify new molecular drivers of lung disease in SSc, as well as how these drivers interact with other genes to influence pathogenesis.

A standard approach to discovering molecular drivers of lung disease in SSc is to identify genetic variants associated with lung outcomes. Genetic studies have been tremendously successful in identifying genetic variants associated with SSc and its complications. In a reflection of the

Copyright © 2020 Tyler *et al.*

doi: <https://doi.org/10.1534/g3.119.400775>

Manuscript received September 26, 2019; accepted for publication October 29, 2019; published Early Online November 6, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.8300543>.

¹Corresponding author: The Jackson Laboratory, Bar Harbor, Maine. E-mail: gregory.carter@jax.org.

complexity of the disease, variants in over 200 genes have been implicated in SSc risk and progression (Yu *et al.* 2010), which has greatly increased our understanding of the development of SSc (Mayes 2012; Agarwal 2010; Agarwal and Reveille 2010) and may aid in personalized disease monitoring and treatment (Assassi *et al.* 2013). The next step in this line of inquiry is to incorporate genetic complexity into models that determine how variants interact with each other to influence disease. By explicitly modeling genetic interactions, or epistasis, we can build understanding of how molecular pathways work in concert to drive SSc pathology.

Initial studies of genetic interactions in SSc have been promising. Epistasis between polymorphisms in the HLA region and cytokines has been shown to predict SSc risk (Beretta *et al.* 2008a), development of severe ventilatory restriction (Beretta *et al.* 2008b), and digital ulcer formation (Beretta *et al.* 2010) in SSc patients. However, progress in this search is limited by a number of challenges. The rarity of the disease and its clinical heterogeneity add to difficulties present in all human genetic studies, such as low minor allele frequencies and the large number of potentially relevant variants. Non-parametric tests such as Multifactor Dimensionality Reduction (MDR) (Hahn *et al.* 2003) have been successful in identifying the interactions that have been identified thus far (Beretta *et al.* 2008a,b 2010). These findings suggest additional, complementary interaction analyses may further dissect the genetic complexity of SSc and other common diseases.

Here we present a novel approach that increases power to detect genetic interactions in human genome-wide association studies (GWAS). We previously developed the Combined Analysis of Pleiotropy and Epistasis (CAPE) to model epistatic interactions in model organisms (Tyler *et al.* 2013; Carter *et al.* 2012). CAPE increases power to detect and interpret genetic interactions by combining information across multiple traits into a single consistent model. We have demonstrated its ability to identify novel genetic interactions not detectable by other methods (Tyler *et al.* 2014, 2016). For this study, we combined CAPE with a filtering step, which filtered the SNPs to those most likely to be involved in genetic interactions. We used Matrix Epistasis (Zhu and Fang 2018), an ultra-fast method for exhaustively testing epistasis in genome-wide SNP data. Candidate SNP pairs were then analyzed with CAPE and significance was assessed with permutation tests.

We applied this approach to genetic and clinical data from a cohort of patients with SSc (dbGaP accession phs000357.v2.p1). To capture aspects of lung disease and autoimmunity, we analyzed two measures of lung function, forced vital capacity (FVC) and diffusion lung capacity (DLC), as well as two autoantibody staining patterns: centromeric, and nucleolar. Anti-centromere autoantibodies (ACA) are associated with pulmonary hypertension, and anti-nucleolar autoantibodies (ANA) are associated with progressive interstitial lung disease and pulmonary arterial hypertension (Betteridge *et al.* 2016). Thus, there may be common genetic pathways underlying both autoantibody status and lung function that we can identify by analyzing all four traits simultaneously.

MATERIALS AND METHODS

Methods workflow

All code used for computational analyses in this paper are available in a supplemental workflow available at https://github.com/annaLtyler/SSc_Epistasis_Workflow.

Genetic and clinical data

We obtained genotype and phenotype from the database of Genotypes and Phenotypes (dbGaP) Accession phs000357.v1.p1.

The full data comprised 833 SSc patients, including 741 females and 85 males. Genotypes were measured at 601,273 SNPs using the Human610_QuadV1_B platform. Patients were assessed for the presence of multiple autoantibodies and lung function tests were performed. Values for autoantibody phenotypes were 1 (present) and 0 (absent), and values for test of lung function included forced vital capacity and diffusion lung capacity. Both measurements were represented as the percent of the predicted value based on factors such as age and sex (Agarwal *et al.* 2009).

We analyzed two lung function traits, percent predicted forced vital capacity (FVCP) and percent predicted diffusion lung capacity (DLCP) as well as two autoantibody patterns, anti-centromere (ACA) and anti-nucleolar (ANA). There were 416 patients (369 females and 47 males) with measurements for all four traits, and all subsequent analyses used only these patients.

Expression data

We obtained gene expression data from the Gene Expression Omnibus (GEO) (Edgar *et al.* 2002; Barrett *et al.* 2013). The data set we used (accession number GSE76808) compared gene expression in lung biopsies taken from patients with SSc and interstitial lung disease (ILD) to biopsies of unaffected tissue from lung cancer patients undergoing surgical resection (Christmann *et al.* 2014). This is the only data set available that compares SSc lung to normal lung. The available data were \log_2 normalized, thus we did not perform further normalization. Group-wise comparisons were performed with two-tailed Student's two-sample *t*-tests.

SNP filtering

We first reduced to 243,662 SNPs with minor allele frequency (MAF) ≥ 0.1 , a relatively high cutoff due to our pair-wise testing strategy. Because analysis of all pairwise combinations of the filtered SNPs was computationally infeasible, we further filtered the SNPs to 1500 SNPs that were likely to participate in genetic interactions influencing the four traits. The number 1500 was selected such that the CAPE pipeline would take about 24 hr to run. For this filter we used Matrix Epistasis (Zhu and Fang 2018), an ultra-fast method of calculating interactions between SNPs. Using Matrix Epistasis, we were able to calculate interaction scores for all SNP pairs exhaustively for each of the four traits. We used a permissive *p* value significance threshold of $p < 1 \times 10^6$ to limit the false discovery inflation cause by trait-based SNP selection. We then used an iterative scheme to identify SNPs that were epistatic and pleiotropic (affecting more than one trait) based on this significance threshold. In the first stage of the search, we identified SNP pairs that interacted significantly ($p < 1 \times 10^{-6}$) to influence more than one trait. There were 52 pairs of SNPs that interacted to affect both lung traits. No other SNP pairs influenced more than one trait. In the next stage of the search, we looked for individual SNPs that were in epistatic interactions affecting more than one trait, regardless of the interacting partner. For example, SNP1 might interact with SNP2 to influence ANA, but interact with SNP3 to influence ACA. SNP1 affects both traits through interactions, but by interacting with different partners. In this case, we keep SNP1, but discard SNPs 2 and 3. We searched in this way for SNPs that influenced all four traits as part of interactions, and then all three traits as part of interactions, and so on until we had selected 1500 SNPs that were potentially epistatic and pleiotropic. For this filtering step as well as for CAPE, we used a dominant coding for the SNPs; *i.e.*, loci homozygous for the major allele were coded as 0 and loci that were either heterozygous or homozygous for the minor allele were coded as 1. This coding allows for maximum representation of each two-locus genotype combination, which increases power to detect

dominant effects, but reduces power to detect recessive or additive effects.

Combined Analysis of Pleiotropy and Epistasis (CAPE)

We identified genetic interactions between SNPs using Combined Analysis of Pleiotropy and Epistasis (CAPE) (Tyler *et al.* 2013; Carter *et al.* 2012). This method combines information across multiple traits to infer directed genetic interactions. Combining information across traits not only increases power to detect genetic interactions, but also allows inference of the allele-to-allele direction of the interaction, which can be either positive (enhancing) or negative (suppressing). We have applied this analysis to multiple model organisms including yeast (Carter *et al.* 2012), fly cell lines (Carter 2013), and mice (Philip *et al.* 2014; Tyler *et al.* 2014, 2017, 2016). This is the first application to a human cohort.

As a preliminary analysis, we tested whether SNPs previously associated with SSc could be recapitulated in this small cohort. We performed linear regression to associate each SNP with each of the four traits.

We then decomposed the autoantibody and lung traits to four orthogonal eigentraits (ETs) using singular value decomposition (SVD) (Figure 2). This step concentrates trait variance into composite ETs potentially improving the ability to map weak genetic effects that are otherwise distributed across traits. We analyzed the first two ETs, which explained 71% of the variance across the four phenotypes (Figure 2A). These ETs contrasted each of the autoantibodies with the other autoantibody and the two lung traits in turn (Figure 2A).

We performed linear regression on SNP pairs and each ET. To avoid testing SNPs in linkage disequilibrium, we did not test any SNP pairs whose Pearson correlation coefficient exceeded an arbitrary, stringent cutoff of 0.5. Out of 1,125,750 possible pairs between the 1500 SNPs plus the covariate sex, 1,125,593 SNP pairs (> 99.9%) passed this criterion. We also assessed population structure in the patient cohort. Population structure, caused by heterogeneous relatedness among individuals in a cohort, can artificially inflate the association between SNPs and traits. To test whether population structure was influencing our association tests, we compared the quantile normalized distribution of the $-\log_{10}(p)$ from association tests with each trait to the distribution of quantile normalized p values under the null hypothesis (Fig S1). There was no systematic inflation of the estimates of significance. We therefore did not correct for population structure in this study.

After performing all pairwise regressions, we recast the regression coefficients to CAPE parameters describing allele-to-allele influences. This process is described in detail elsewhere (Carter *et al.* 2012; Tyler *et al.* 2013), and we describe it briefly here. The first step in calculating CAPE coefficients is a reparametrization of the regression coefficients to obtain delta parameters (δ_1 and δ_2) which describe the degree to which the presence of one SNP influences the phenotypic effects of the other SNP:

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \beta_1^1 & \beta_2^1 \\ \beta_1^2 & \beta_2^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \beta_{12}^1 \\ \beta_{12}^2 \end{bmatrix}$$

These parameters are then translated into directional (m_{12} and m_{21}) coefficients that self-consistently describe how the two SNPs influence one another:

$$\delta_1 = m_{12}(1 + \delta_2), \delta_2 = m_{21}(1 + \delta_1)$$

We estimated standard errors for these interaction terms by propagating standard error terms from least squares regression with

second-order Taylor expansion on the regression parameters (Carter *et al.* 2012).

We estimated significance of the resulting model parameters through permutations with family-wise error rate estimation. We generated a null distribution as follows: ETs were randomly re-ordered relative to the genotypes, and we performed the single-SNP scan, and the pairwise scan as with the original data. We repeated this process until we had generated a null distribution with 1.5 million values. We corrected empirical p values using false discovery rate (FDR) (Benjamini and Hochberg 1995) and used a significance threshold of $q < 0.05$.

To aid in visualizing the strength of each interaction we calculated a predicted additive effect of each pairwise SNP combination. These are shown in figure 2. We used the linear regression coefficients as shown in the following model.

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_1x_2\beta_{1,2} + \epsilon$$

We subtracted the intercept (β_0) from each term, and calculated the predicted additive effect as the sum of the two main effect β coefficients. Similarly, we defined the error as the sum of the standard errors of each group defined by the pairwise genotypes.

Assignment of SNPs to genes

To assign each SNP in the final network to a gene, we downloaded SNP annotations from SNP Nexus (Dayem Ullah *et al.* 2018, 2013, 2012) and assigned each gene to the nearest or containing gene.

Coexpression analysis

To explore the potential function of the interactions we identified, we used the Search-Based Exploration of Expression Compendium (SEEK) (Zhu *et al.* 2015). SEEK is a web tool (<http://seek.princeton.edu>) that searches across thousands of gene expression data sets to find genes that are co-expressed with a user-defined query gene or gene set. The search can be restricted to data sets in individual tissues or that have associated keywords, thus providing context for the co-regulation.

Based on the CAPE results, we analyzed two gene pairs: *WNT5A* and *RBMS3* in one query and *RBMS3* and *MSI2* in another. Using SEEK, we searched for co-regulated genes across all expression data sets. We also restricted our search to subsets of data sets to identify co-regulated genes in particular contexts (Table 2). For example, to look for tissue-specific co-regulation, we searched for genes co-regulated in all lung tissue data sets and also in all skin tissue data sets. Both tissues are relevant to SSc pathogenesis. However, in this study, we identified genetic interactions associated with lung function. We thus expected that genes co-regulated with the query genes in lung would better reflect the function of each interaction in this study.

In addition to tissue-specific searches, we also searched for genes co-regulated with our query genes in disease-specific data sets. To perform these searches, we restricted the search to data sets associated with key words, for example autoimmunity and fibrosis.

In each of these queries, we identified the genes whose expression was significantly correlated with the pair of query genes ($p < 0.05$). These gene sets represent genes that were co-regulated with the query gene set across multiple tissues and disease contexts. We propose that functional enrichments of these gene sets are related to the function of the query gene pair in this disease cohort. To identify functional enrichment of each gene set, we used gProfileR (Reimand *et al.* 2016) (See workflow).

Among the significantly enriched terms, were many that are known to be related to SSc, for example terms relating to extracellular matrix organization and wound healing. We were interested in determining

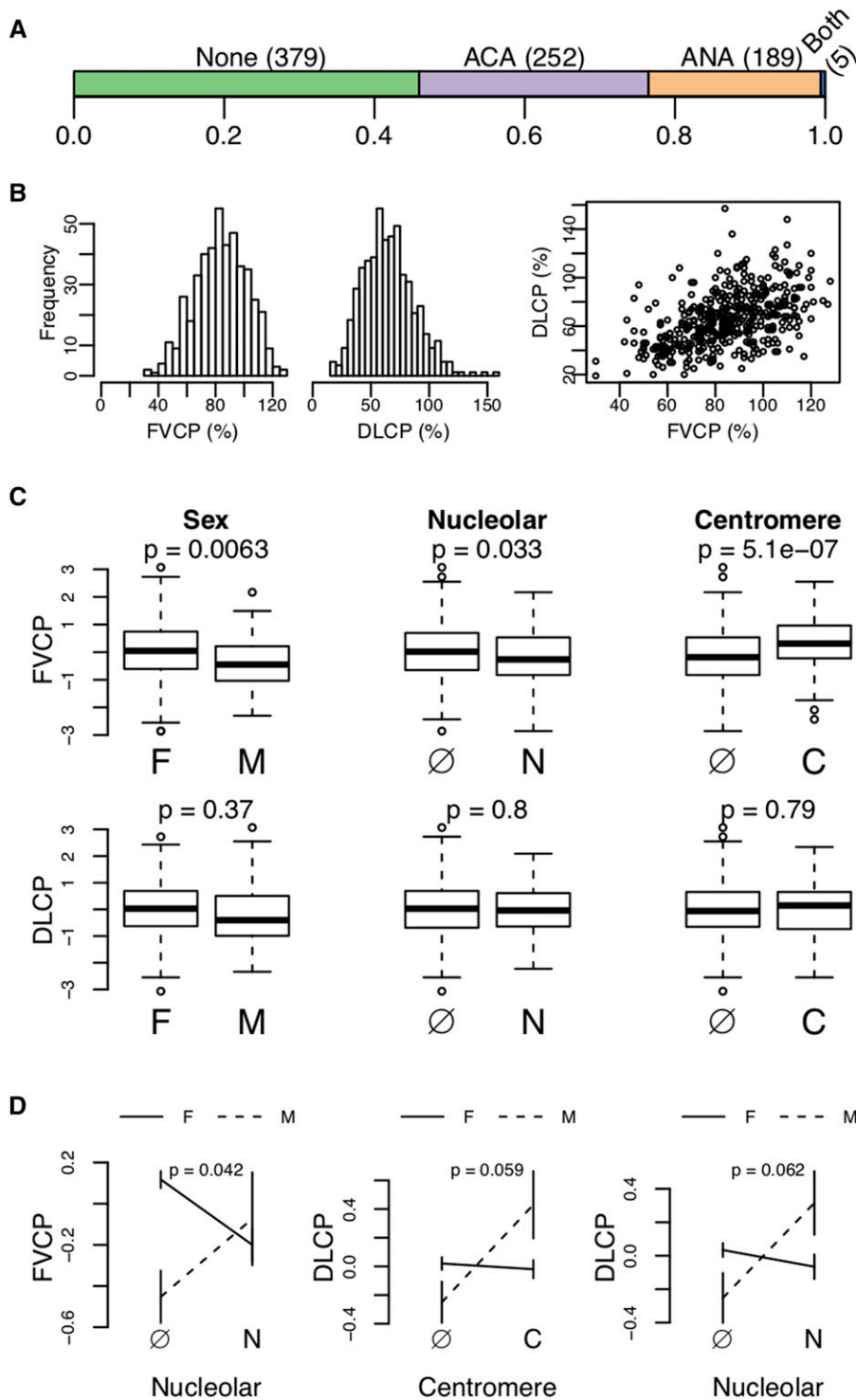


Figure 1 Trait distributions and correlations. (A) Colored bars show the proportion of patients that were negative for both autoantibodies (None), positive for either centromere (ACA) or nucleolar (ANA) autoantibodies, or positive for both (Both). (B) Histograms showing the distribution of FVCP (%) and DLCP (%). The distributions are roughly normal. The third panel shows the correlation between FVCP and DLCP. They were moderately correlated ($r = 0.48$, $p = 1.3 \times 10^{-25}$). (C) Box plots show associations between lung traits and binary traits. Sex, nucleolar status, and centromere status all had significant effects on FVCP. None of the binary traits were significantly associated with DLCP. P values from two-tailed Student's *t*-test ($p < 0.05$) are reported. Autoantibody status is denoted as ANA (N), ACA (C) or absent (\emptyset). (D) These plots show interactions between binary traits that affect the lung traits. Sex and nucleolar status interacted significantly to influence FVCP ($p \leq 0.05$). Interactions between sex and autoantibody status on DLCP were not quite significant, but trended toward significance ($p \approx 0.06$).

whether these sets of enriched GO terms enriched were related to SSc more than we would expect at random. To assess this, we downloaded a set of disease-GO term associations from the Comparative Toxicogenomics Database (<http://ctdbase.org/downloads/#phenotypediseases>) (Davis *et al.* 2019, 2016). This database provided us with a set of GO terms that are associated with SSc.

From the database, we collected all GO terms associated with SSc, including Systemic Scleroderma, Diffuse Scleroderma, and Limited Scleroderma, and looked for overlap between these GO terms and the

GO terms enriched in the SEEK analysis. To determine significance of the overlap between gene pair GO terms and SSc-associated GO terms, we performed Fisher's Exact Test. Significance was assessed at $p \leq 0.05/16$.

For visualization purposes, we clustered SSc-related GO terms using REVIGO (Supek *et al.* 2011) (<http://revigo.irb.hr>).

Interestingly, although Wnt signaling has been shown to be involved in SSc, no Wnt signaling GO terms were associated with SSc by CTD. This indicates missing true associations between SSc and GO terms and suggests that the associations we calculated are conservative estimates.

■ **Table 1** Information about position of SNPs significantly associated with ACA ($p < 0.01$)

SNP	Chr	Position	Nearest Gene	Annot.	Dist. (bp)
rs4248166	6	32398644	BTNL2	intronic	
rs660895	6	32609603	HLA-DRB1	upstream	19755
rs7755224	6	32684540	HLA-DQB1	upstream	16157
rs9275390	6	32701379	HLA-DQB1	upstream	32996

Data availability

Genotype and phenotype data are available from the Database of Genotypes and Phenotypes (DbGaP) (<https://www.ncbi.nlm.nih.gov/gap>) Accession: phs000357.v2.p1. Gene expression data are available from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) Accession: GSE76808 A complete workflow including all code to generate results and figures is available here: https://github.com/annaLtyler/SSc_Epistasis_Workflow. Supplemental material available at figshare: <https://doi.org/10.25387/g3.8300543>.

RESULTS

Trait descriptions

We analyzed a cohort of 416 patients with lung function and autoantibody measurements (369 females and 47 males). The two lung traits, forced vital capacity (FVCP) and diffusion lung capacity (DLCP), were measured as a percentage of the total predicted by patient demographic parameters, such as age and weight (Agarwal *et al.* 2009). The autoantibodies were distributed as shown in Figure 1A. Roughly half the patients were negative for both antinucleolar autoantibodies (ANA) and anti-centromeric autoantibodies (ACA), and the other half of patients were positive for one or the other. Only five patients presented with both autoantibodies.

Both lung traits were roughly normally distributed, with maximum values greater than 100%, and the traits were modestly correlated (Pearson's $r = 0.48$, $p = 1.3 \times 10^{-25}$) (Figure 1B). We used rank Z normalization on both traits prior to performing association tests.

FVCP was influenced by sex and autoantibody status (Figure 1C). Males had significantly lower FVCP on average than females (two-tailed Student's t -test $p = 0.0063$). The presence of the two autoantibodies had opposite effects. The presence of ANA significantly reduced FVCP (two-tailed Student's t -test $p = 0.033$), while the presence of ACA significantly improved FVCP (two-tailed Student's t -test $p = 5.1 \times 10^{-7}$). Neither autoantibody nor sex had any significant association with DLCP (all two-tailed Student's t -test $p > 0.05$).

We also looked for interactions between the autoantibodies and sex that influenced the lung traits (Figure 1D). Presence of ANA and sex interacted nominally significantly to influence FVCP ($p = 0.042$). In this interaction, females positive for ANA had lower FVCP on average than females negative for ANA. However, in males, the effect was the opposite. Males positive for ANA had higher FVCP on average than males negative for ANA. None of the other factor combinations had any significant effects on either FVCP or DLCP (all two-tailed Student's t -test $p > 0.05$); however sex and autoantibody status trended toward a significant interaction for DLCP. The presence of each autoantibody had very little effect on DLCP in females, but increased DLCP in males.

Single-locus association identified four SNPs with significant main effects on ACA

The patients in this study were genotyped at 601,273 SNPs. However, because detection of genetic interactions can be confounded by low minor allele frequencies (MAF), we first filtered this data set to 243,662 SNPs with $MAF \geq 0.1$. We then performed all single marker

regressions (Materials and Methods). The purpose of this analysis was twofold. First it allowed us to determine whether the number of patients in this cohort was sufficient to recapitulate known associations between SNPs and SSc traits. Second, we used the p value distributions from this analysis to investigate whether population structure in this cohort might affect the SNP associations (Materials and Methods). There was no systematic effect of population structure in this cohort (Fig S1). We therefore did not use a correction for population structure in this analysis.

The SNP association tests identified significant main effects in the HLA region influencing the presence of ACA (Table 1). One of these SNPs (rs9275390) maps upstream of HLA-DBQ1 and was previously associated with positive ACA status in an SSc GWAS (Gorlova *et al.* 2011). Another SNP, rs660895, maps upstream of HLA-DRB1, and was previously associated with risk of the autoimmune disease Rheumatoid Arthritis (Plenge *et al.* 2007).

Multiple epistatic interactions influenced lung and autoantibody traits

We next performed pairwise SNP associations to identify interactions between SNPs. As CAPE is relatively computationally intensive, performing exhaustive pairwise testing ($> 1.8 \times 10^{11}$ tests) was not feasible. We thus filtered the SNPs to 1500 that were likely to participate in epistatic interactions influencing more than one trait (Materials and Methods). This number was selected such that the CAPE pipeline could be completed in about 24 hr. In addition, we decomposed the traits into orthogonal eigentraits (ETs) using singular value decomposition (SVD) (Figure 2A) (Materials and Methods). Traits that are moderately correlated may share common underlying biological processes. Using ETs concentrates signals from these underlying processes into single traits thereby increasing power to map them genetically. We selected the first two ETs for analysis. Each ET describes the contrast between one autoantibody and the two lung traits. After calculating CAPE coefficients associated with the two ETs, we rotated the coefficients back to trait space thus identifying interactions associated with all traits simultaneously. By selecting only two ETs we lost some biological information contained in the other two ETs, but we simultaneously gained power to identify interactions influencing the biological processes encoded by the first two eigentraits.

We identified a network of 11 significant genetic interactions involving 15 SNPs and the covariate sex (Figure 2B) ($FDR \leq 0.05$). The two most significant interactions (largest standardized effects) formed a single subnetwork among SNPs rs556874, rs1449292, and rs792399 (Figure 2B). This subnetwork was notable in that, not only did it contain the most significant interactions, but all three SNPs were intronic in protein coding genes *WNT5A*, *RBMS3*, and *MSI2*. The SNP rs556874 is located in the fourth intron of *WNT5A*, which is a member of the large family of Wnt ligands. Wnt signaling is a widely important family of signaling pathways integral to embryonic development, carcinogenesis, and many other processes. The SNP rs1449292 is located in an intron of *RBMS3*. This gene encodes a member of a small family of proteins that bind single-stranded RNA and DNA to regulate a wide range of biological processes. The SNP rs792399 is located in an intron of *MSI2*

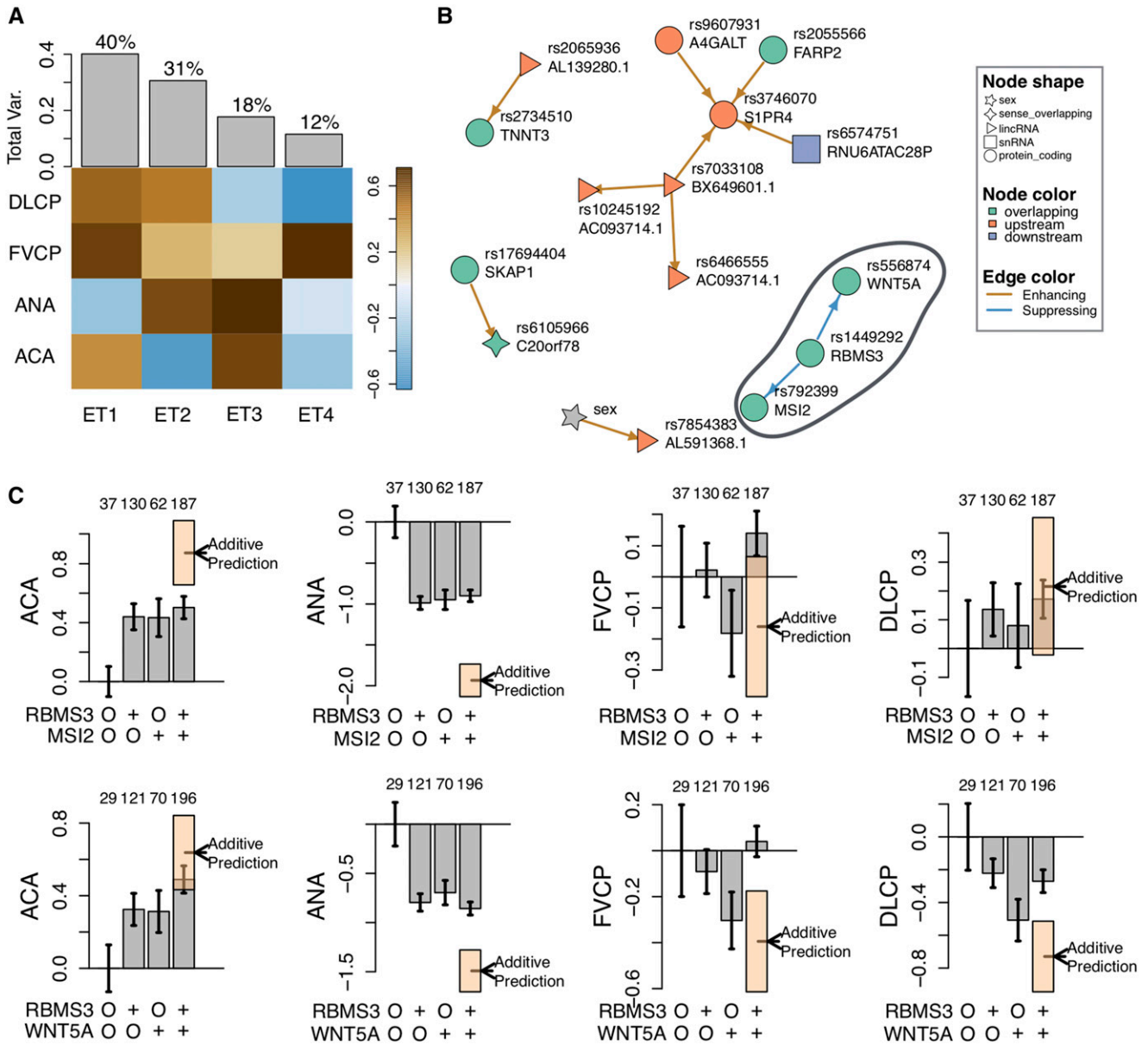


Figure 2 Genetic interactions influenced lung function and autoantibodies in SSc. (A) The relative contributions of each trait to each eigentrait (ET). Gray bars indicate how much overall variance each ET captures. We used the first two ETs, each of which contrasted one of the autoantibodies, with the other autoantibody and the lung function traits. These two ETs captured 71% of the overall trait variance. (B) The network of significant SNP-SNP interactions. Each node in the network represents one SNP. Each is labeled with the rs number, as well as with the name of the nearest gene. The color of each node indicates whether the SNP is upstream, downstream, or overlapping the gene. The shape of each node indicates the type of gene, protein coding, lncRNA, etc. The links between the nodes are colored to indicate whether each significant interaction was enhancing or suppressing. Three SNPs (rs556874, rs1449292, and rs792399) formed a three-node subnetwork we called the Wnt subnetwork (circled). Each of these SNPs overlaps a protein coding gene. (C) The effects of the Wnt subnetwork interactions on clinical traits. The x-axis in each plot indicates the four possible genotypes for each set of two SNPs. SNPs are labeled by their overlapping gene for clarity. The “O” indicates the reference genotype, and “+” indicates the alternate genotype. Gray bars indicate the mean trait value for the group of patients with the genotype indicated on the x-axis. The number above each bar indicates how many patients are in that group. Error bars show standard error. The additive prediction for each double alternate genotype is indicated by the dashed line and the error of the estimate is shown in the orange box. Gray bars that fall outside the orange box indicate interactions between the alleles. In A and C traits are abbreviated as follows: percent predicted forced vital capacity (FVCP) and percent predicted diffusion lung capacity (DLCP), antinucleolar autoantibodies (ANA), anticentromere autoantibodies (ACA).

about which little is known. Although both *WNT5A* and *RBMS3* are both encoded on Chr 3, there is no LD between the interacting SNPs in this cohort ($R^2 = 0.004$). We focused on this subnetwork, which we refer to as the “Wnt subnetwork,” for the remainder of the analysis.

Interactions in the Wnt subnetwork were less than additive

The interactions in the Wnt subnetwork were suppressing (Figure 2B). This means that in the presence of the alternate allele at one locus, the

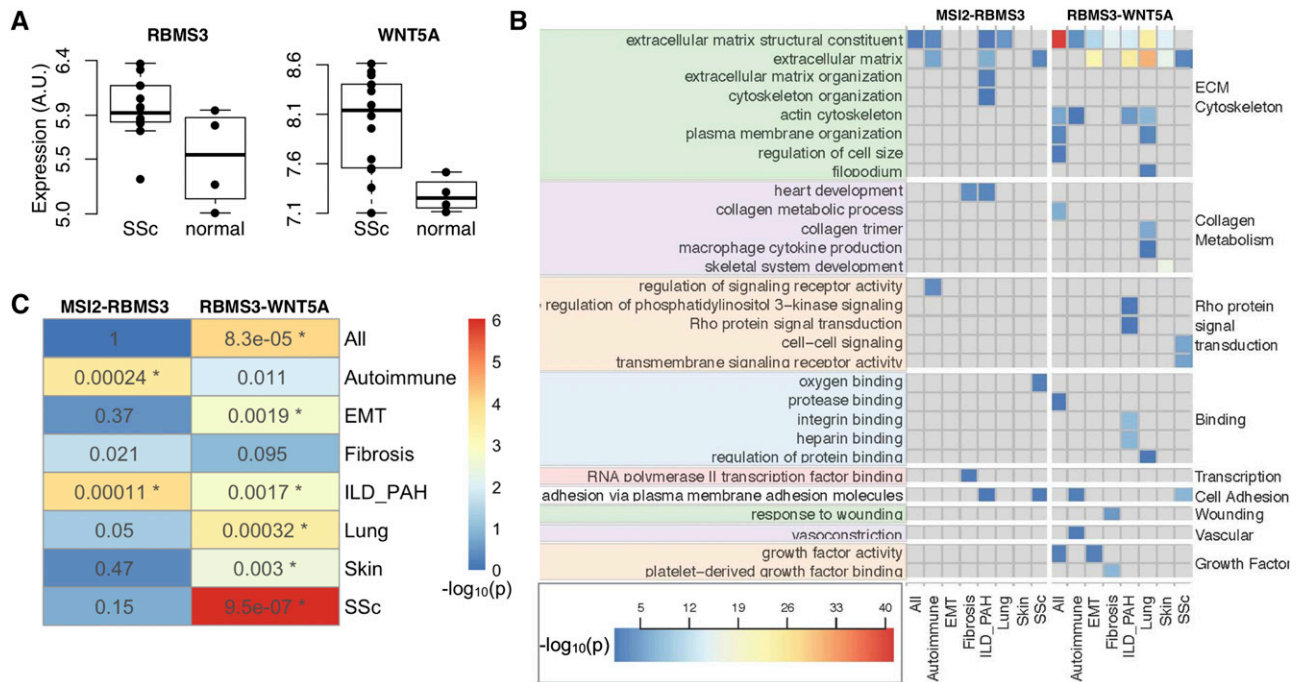


Figure 3 Expression of subnetwork genes in SSc. (A) Both *WNT5A* (two-tailed Student's *t*-test $p = 0.011$) and *RBMS3* (two-tailed Student's *t*-test $p = 0.02$) were significantly overexpressed in lung tissue from SSc patients with interstitial lung disease. (B) The $-\log_{10}(p)$ for enriched GO terms in gene sets derived from SEEK (Materials and Methods). Each column shows results from a subset of the data sets on GEO. They are as follows: All - all data sets on GEO, Autoimmune - data sets studying autoimmunity, EMT - data sets studying epithelial to mesenchymal transition, Fibrosis - data sets studying fibrosis, ILD_PAH - data sets studying interstitial lung disease (ILD) or pulmonary arterial hypertension (PAH), Lung - data sets using lung tissue, Skin - data sets using skin tissue, SSc - data sets studying SSc. Each row shows the results for one GO term identified on the right-hand side. The GO terms are grouped functionally into nine groups identified by the GO term clustering algorithm REVIGO (Supek et al. 2011): ECM/cytoskeleton, collagen metabolism, Rho protein signaling transduction, Binding, Transcription, Cell Adhesion, Wound Repair, Vascular Function, and Growth Factors. The gene names at the top of the figure indicate that the left hand side of the figure shows results for the *RBMS3*-*MSI2* interaction, and the right side shows results for the *RBMS3*-*WNT5A* interaction. The legend indicates how the colors of the cells relate to the $-\log_{10}(p)$ of the term enrichment. Only GO terms with fewer than 300 genes are shown. (C) Significance of overlap between gene pair-enriched GO terms and SSc-related GO terms. Fisher's exact test *p* values are reported for each condition (rows) and each gene pair (columns). Asterisks indicate $p \leq 0.5/16$. Colors denote the $-\log_{10}(p)$.

phenotypic effects of the interacting locus were suppressed. In this case, the suppressing interactions resulted less-than-additive interactions. These effects can be seen in Figure 2C. The *RBMS3*-*MSI2* interaction primarily affected the two autoantibody traits. The alternate alleles of the SNPs each had a negative effect on autoantibody presence. However, the effects of both alternate alleles together was redundant and did not decrease autoantibody presence beyond either of the individual alleles. The effects of this interaction on lung function were not different from additive.

The *RBMS3*-*WNT5A* interaction affected both lung function traits and ANA. The effects on ANA were similar to the effects of the *RBMS3*-*MSI2* interaction: individual alternate alleles decreased ANA incidence, but having two alternate alleles did not further decrease incidence beyond that seen for a single alternate allele. For both lung function traits, the *WNT5A* alternate allele had a large negative effect in the presence of the *RBMS3* reference allele, but this effect was completely suppressed in the presence of the *RBMS3* alternate allele.

Expression of *RBMS3* and *WNT5A* was dysregulated in SSc lung

We investigated whether the candidate genes in the Wnt subnetwork were differentially expressed in SSc patient lung tissue (Materials and Methods) compared to control lung tissue. Differential expression between SSc lung tissue and control lung tissue suggests potential

involvement in SSc pathogenesis. *WNT5A* (two-tailed Student's *t*-test $p = 0.02$) and *RBMS3* (two-tailed Student's *t*-test $p = 0.011$) had significantly higher expression in SSc lung tissue relative to healthy lung tissue (Figure 3A). Although we cannot assess the role of the genetic variants in expression level, differential expression supports the hypothesis that *RBMS3* and *WNT5A* may be involved in SSc processes. We could not assess the differential expression of *MSI2*, as it was not present in the data set.

Co-expressed genes were functionally enriched for processes related to SSc

To further investigate biological processes affected by the Wnt subnetwork, we searched across GEO for genes that were co-regulated with the gene pairs in the Wnt subnetwork. Genes with correlated expression often function together, and co-expression is frequently used to elucidate gene function (Eisen et al. 1998; Whitfield et al. 2002; Langfelder and Horvath 2008; Mahoney et al. 2015). We used this principle to identify possible functional roles of the interactions in the Wnt subnetwork. To investigate whether the interactions might be context-specific, we searched across multiple subsets of data sets in GEO (Materials and Methods). For example, to identify lung-specific functions of each gene pair, we searched for genes with correlated expression among data sets that were obtained from lung tissue. To identify functions that were related to autoimmunity,

■ **Table 2** The terms used to refine searches across gene expression data sets in GEO. Subsets were defined either by selecting a pre-defined subset (categorical) or by using key words (keyword) (Materials and Methods). The final column indicates how many data sets were included when subset criteria were applied

Subset	Method	Data sets
All	none	3402
Autoimmune	categorical	18
EMT	keyword	13
Fibrosis	categorical	39
ILD_PAH	keyword	4
Lung	categorical	239
Skin	categorical	136
SSc	keyword	6

we searched only among data sets that analyzed autoimmune disease. All data set subsets are listed in (Table 2).

In each of these data set subsets, we queried the *WNT5A-RBMS3* gene pair separately from the *RBMS3-MSI2* gene pair. For each query, we identified sets of genes whose expression was significantly correlated with the two query genes ($p < 0.05$), and used the R package gProfileR (Reimand *et al.* 2016) to identify functionally enriched GO terms in each gene set. Multiple GO terms were significantly associated with the gene set from each query. For a complete list of the enriched terms, see Fig S2. To focus on terms that are strongly related to SSc, we used disease-GO term associations provided by the Comparative Toxicogenomic Database (Davis *et al.* 2019, 2016). These terms, and their enrichment p values across all queries are shown in Figure 3B.

Genes coexpressed with *RBMS3* and *WNT5A* were functionally enriched for multiple relevant processes, including ECM organization, wound repair, vasoconstriction, and growth factor activity.

Both gene pairs were significantly correlated with extracellular matrix (ECM) GO terms across multiple tissue types and conditions (Figure 3); however, only the *RBMS3-WNT5A* pair was significantly correlated with genes specifically involved in collagen metabolism indicating that this pair may affect ECM through effects on collagen, whereas the *RBMS3-MSI2* pair may influence ECM in a collagen-independent manner.

These enrichments for collagen metabolism appeared in lung tissue experiments, but not skin suggesting that the gene pair may be involved in collagen metabolism in a tissue-specific manner. Similarly, most significant enrichments associated with the *RBMS3-MSI2* interaction were concentrated in lung, lung disease, and autoimmune conditions, suggesting that this interaction may be specific to lung tissue, and may be aberrantly regulated in autoimmune conditions.

Fisher's exact tests showed that many of the groups of GO terms enriched in genes co-expressed with the epistatic gene pairs were significantly related to SSc processes (Figure 3C). Genes co-expressed with the *RBMS3-WNT5A* pair were significantly SSc-related across more conditions than the genes co-expressed with *RBMS3* and *MSI2*. The patterns for the two gene pairs suggest that the *RBMS3-MSI2* interaction may be more involved in autoimmunity, whereas, the *RBMS3-WNT5A* interaction may be more involved specifically in lung disease and SSc-specific processes.

Positive interaction also highlights promising candidates

Although we focus here primarily on the Wnt subnetwork, we also investigated whether any of the enhancing interactions in the genetic

interaction network showed promise for follow-up. The most significant positive interaction in the network involved two additional variants located within or very near protein coding genes. This interaction was between rs2055566, which is located in an intron of *FARP2*, and rs3746070, which is located 736 bp upstream of *S1PR4*. These two variants had additive effects on all traits except DLCP. Neither variant had a significant main effect on DLCP, but individuals with both alternate alleles, had improved DLCP relative to other individuals (Fig S3).

DISCUSSION

We used a novel analytical approach to identify an epistatic interaction network influencing lung function in patients with SSc. In this approach we used three strategies to increase power to detect interactions in a patient cohort. First, we used a SNP filtering step to select the most highly epistatic and pleiotropic SNPs for further testing. This step reduced the number of statistical tests performed, but may also increase false positive rate due to the trait-based nature of the selection. We addressed this in part by using permutation-based p values and a stringent FDR cutoff (≤ 0.05). We also noted that none of the interactions identified by CAPE overlapped with the interactions identified by Matrix Epistasis, *i.e.*, all CAPE interactions were novel combinations of SNPs selected by the filtering step; however, the results reported are intended for discovery, and validation is required.

As a second strategy for increasing power to detect interactions, we combined information across multiple related traits: two auto-antibody traits, and two lung function traits. And third, we decomposed the traits into orthogonal eigentraits and used only the top two eigentraits in our study. This step reduced noise and captured mappable signals that related to all four traits. Thus, using Matrix Epistasis to pre-filter SNPs combined with CAPE allowed us to identify genetic interactions in a relatively small human cohort. Because of the increased false positive rate imposed by the SNP filtering step, the results reported here are primarily for discovery and hypothesis generation. Although the genes identified are supported by the literature as discussed below, the interactions between them need to be replicated in an independent cohort.

The resulting genetic interaction network contained a notable sub-network consisting of two interactions between three SNPs. Because the interactions in this network were highly significant and because all three SNPs were located within protein-coding gene bodies, we focused on this subnetwork for further analysis. We also briefly investigated the most significant positive interaction between SNPs that were also associated protein coding genes, *FARP2* and *S1PR4*. Each of the candidates implicated by these interactions has been associated either directly with SSc or with SSc-related processes, which speaks to the reliability of this analysis.

The two genes participating in the positive interaction have been more directly linked to SSc than the genes in the Wnt subnetwork. *S1PR4*, for example, encodes the sphingosine-1-phosphate receptor 4, and sphingosine-1-phosphate well is known to affect SSc pathogenesis. This signaling phospholipid is elevated in the serum of SSc individuals and is capable of producing many of the abnormalities seen in SSc (Pattanaik and Postlethwaite 2010; Tokumura *et al.* 2009). Sphingosine-1-Phosphate Receptor 5 was found to modulate early fibrogenesis in a mouse model of SSc (Schmidt *et al.* 2017), and the selective *S1P1* receptor modulator cenerimod has been shown to attenuate lung and skin fibrosis in two different mouse models of SSc (Kano *et al.* 2019). *FARP2* encodes FERM, RhoGEF and pleckstrin domain protein 2 and is involved in semaphorin signaling. Dysregulation of semaphorin signaling has been associated with antibody production, disease type,

thickening of skin, disease duration, and active inflammation in SSc patients (Besliu *et al.* 2011; Rimar *et al.* 2015).

The genes in the three-gene Wnt subnetwork also have support in the literature for SSc involvement. We focus the discussion primarily on these genes to highlight the possible connections between genetic interactions and functional relatedness. Although none of the genes in the Wnt subnetwork had previously been associated with lung function in SSc, they have all been implicated in lung disease, as well as in molecular processes that are known to be involved in SSc pathogenesis.

WNT5A, and Wnt signaling more generally, have been implicated in many processes known to be involved in SSc pathogenesis (Tsou and Sawalha 2017; Wei *et al.* 2011; Katsumoto *et al.* 2011; Lafyatis 2012), such as angiogenesis (Huang *et al.* 2005), keratinocyte differentiation and inflammation signaling (Wang *et al.* 2018), and fibroblast proliferation (Vuga *et al.* 2009). *WNT5A* has also been linked to cellular transdifferentiation, a family of processes by which differentiated cell types dedifferentiate and redifferentiate into another cell type. Transdifferentiation processes are critical for embryonic development, wound healing and tissue regeneration (Kalluri and Weinberg 2009). However, when they are dysregulated, these processes contribute to metastases in cancer, and widespread fibrosis seen in fibrotic diseases (Kalluri and Weinberg 2009). There are three types of transdifferentiation that have been linked to fibrosis in SSc: epithelial to mesenchymal transition (EMT) (Postlethwaite *et al.* 2004), endothelial to mesenchymal transition (EndoMT) (Manetti *et al.* 2017) and adipocyte-myofibroblast transition (AMT) (Marangoni *et al.* 2015).

Each of these processes has been hypothesized to be a source of excess myofibroblasts in SSc tissues. Myofibroblasts differentiate from fibroblasts in response to signaling from TGF- β , as well as other growth factors and cytokines. In healthy individuals, myofibroblasts play an important role in wound healing. They contract to close wounds and deposit ECM to seal wounds and to provide a scaffold for the re-establishment of epithelial tissue. In SSc and other fibrotic disease however, there are increased numbers of myofibroblasts with a concomitant increase in the deposition of extracellular matrix, which causes tissue stiffening (Postlethwaite *et al.* 2004; Nakamura and Tokura 2011). EMT has been directly implicated in lung fibrosis in mice (Kim *et al.* 2006; Hashimoto *et al.* 2004), and may be an important source of myofibroblasts in SSc skin (Nakamura and Tokura 2011). Wnt signaling, and *WNT5A* in particular, has been shown to mediate EMT in melanoma (Dissanayake *et al.* 2007) and in lung cancer (Wang *et al.* 2017).

Endothelial to mesenchymal transition (EndoMT) has also been proposed to be a source of myofibroblasts in SSc (Piera-Velazquez *et al.* 2011; Manetti *et al.* 2017; Jimenez and Piera-Velazquez 2016). Originally thought only to occur in developing embryos, EndoMT has been demonstrated to occur in adult cattle (Arciniegas *et al.* 1992) and mice (Zeisberg *et al.* 2007). TGF- β -induced EndoMT has since been shown to be involved in the bleomycin model of pulmonary fibrosis as well as SSc-associated pulmonary hypertension (Jimenez 2013; Jimenez and Piera-Velazquez 2016). Furthermore, there is evidence to suggest that Wnt signaling may mediate EndoMT (Jimenez and Piera-Velazquez 2016).

The third transdifferentiation process to be proposed as a source of myofibroblasts in SSc is AMT (Marangoni *et al.* 2015). Marangoni *et al.* (2015) showed that when adipose-derived stem cells from healthy donors were incubated with TGF- β , they lost markers of adipocytes, and gained markers of myofibroblasts. In this experiment *WNT5A* was strongly upregulated in the transitioning cells.

Like *WNT5A*, *RBMS3* has been linked to transdifferentiation. Upregulation of *RBMS3* inhibits EMT in lung cancer through the inhibition of canonical Wnt signaling (Yang *et al.* 2018; Liang *et al.* 2015). By downregulating MMP2 and β -catenin, *RBMS3* also inhibits microvessel formation (Chen *et al.* 2012). Loss of microvessels and reduced angiogenesis are hallmarks of SSc (Manetti *et al.* 2010; Distler *et al.* 2004). *RBMS3* may also directly influence collagen synthesis (Penkov *et al.* 2000; Fritz and Stefanovic 2007), which is highly upregulated in SSc. Although *RBMS3* has not been directly linked to SSc, SNPs in *RBMS3* have been associated with risk of another autoimmune disease affecting connective tissue called Sjögren's syndrome (Song *et al.* 2016).

Of the three genes in the Wnt subnetwork, the least is known about *MSI2*, and it is most studied in the context of cancer and its influence on EMT (He *et al.* 2014; Kudinov *et al.* 2016). *MSI2* is upregulated in metastatic-competent lung cancer cell lines, and aggressively metastatic patient tumors have upregulated *MSI2* (Kudinov *et al.* 2016). Conversely, depletion of *MSI2* resulted in reduced metastatic potential (Kudinov *et al.* 2016). As mentioned previously, EMT may be an important process in SSc pathogenesis. *MSI2* is also related to Wnt signaling and may influence hepatocellular carcinoma outcomes through dysregulation of Wnt signaling (Wang *et al.* 2015).

Thus, all three genes in the Wnt subnetwork have been implicated in mediating the transdifferentiation process of epithelial to mesenchymal transition (EMT). Both EMT (Willis and Borok 2007) and EndoMT (Jimenez and Piera-Velazquez 2016) are known to take place in the lung and have demonstrated roles in SSc (Jimenez 2013; Postlethwaite *et al.* 2004). These results suggest that EMT, or another similar transdifferentiation, may play a role in lung disease pathogenesis and lung function in SSc patients, and that genes in the Wnt subnetwork we identified here may be involved.

Furthermore, all three genes in the Wnt subnetwork have been shown to be associated with Wnt signaling. These pathways are known to be important both in lung development, and in lung fibrosis in SSc (Pongracz and Stockley 2006; Lam *et al.* 2011). Wnt signaling, both canonical (β -catenin mediated) and non-canonical (not β -catenin mediated), has been shown to drive EMT. Upregulation of *WNT5A* increases EMT in lung cancer (Wang *et al.* 2017) possibly through stimulation of the non-canonical PKC Wnt pathway (Dissanayake *et al.* 2007). Similarly, upregulation of *MSI2* is an indicator of increased metastatic potential in lung cancer (Kudinov *et al.* 2016) and hepatocellular carcinoma (Kudinov *et al.* 2016), and is proposed to stimulate EMT through stimulation of canonical Wnt signaling (Wang *et al.* 2015). *RBMS3* has been shown to reduce EMT and metastatic capacity of lung cancer cells (Liang *et al.* 2015) and reduces EMT in gastric cancer cells through suppression of canonical Wnt signaling (Yang *et al.* 2018). Taken together, these results suggest that the Wnt subnetwork we identified here may mediate cellular transdifferentiation in the SSc lung through regulation of Wnt signaling.

Importantly, the Wnt subnetwork includes not only three novel candidate genes, but also the genetic interactions between them. These interactions suggest functional relatedness, which is broadly supported through connections in the literature with EMT and Wnt signaling. But we can also make more direct connections between the specific genes by using a more agnostic gene expression-based approach. We looked at genes that were co-expressed with each pair of subnetwork genes across multiple different tissue and disease contexts, including lung tissue, lung disease, and skin tissue. We found that the gene pairs were associated with multiple processes related to SSc pathogenesis, such as ECM organization, growth factor activity, and vascular abnormalities. By comparing the enrichments related to the two interactions, we can generate hypotheses about their specific involvement. For example, both

gene pairs were co-regulated with genes involved in ECM-related GO processes; however only genes co-expressed with *RBMS3* and *WNT5A* were associated with collagen metabolism, suggesting that *RBMS3* and *MSI2* together may influence a distinct aspect of ECM.

Interestingly, functional enrichment of co-expressed genes also appeared to be context-specific. SSc-related functional enrichments were seen predominantly in the lung disease (ILD_PAH) data sets, as well as in autoimmunity data sets, as opposed to skin. We interpret this to mean that both genetic interactions may be tissue- and disease-specific. It is possible that these interactions were only detectable because we looked specifically at lung function outcomes in an SSc cohort. Thus, the interactions we identified may not be related to skin pathology. Validation experiments, therefore, should focus on lung tissue.

We further hypothesize that these interactions may be particularly important to lung function in a disease state. All three SNPs we identified have high minor allele frequencies in the general population (rs556874 MAF \approx 0.4, rs1449292 MAF \approx 0.5, and rs792399 MAF \approx 0.4) (Sherry *et al.* 2001). We hypothesize that these SNPs are unlikely to have any measurable effect on lung function in healthy individuals, but that the interactions between them become important in disease. We further hypothesize that because these alleles are so common, they may not increase the risk of developing SSc, but that they affect the course of the disease once established. We performed our genetic analysis on a cohort of patients with SSc that did not include controls. Although a cohort containing only affected individuals may have few individuals, thereby reducing power to detect interactions, we assert that this type of case-only study is important in detecting disease-specific interactions that would not be detectable in a data set that included controls.

These interactions may be important considerations in personalized medicine. Whereas genetic main effects are averaged across all genetic backgrounds, genetic interactions may identify clinically relevant subgroups of patients based on allele combinations. For example, in this study we found that the minor allele of rs556874 in *WNT5A* had substantially reduced lung function, but only in patients with the major allele of rs1449292 in *RBMS3*. With this knowledge, this subset of patients can be monitored more closely for lung disease and perhaps provided with treatments for lung disease earlier in the course of their disease.

In this study we have taken a systems approach to identifying potential SSc-related processes by analyzing genetic interactions that influence multiple disease traits simultaneously. By addressing the complexity of this disease in our models we increased power to detect novel interactions among highly promising candidate genes. The interactions, furthermore, may be disease- and tissue-specific. Identifying such specific interactions may provide a promising avenue forward in drug target discovery. The genes identified in this study are all widely expressed, and targeting them broadly may result in unintended effects. However, because evidence suggests that the interactions between the genes may be tissue- and disease-specific, perhaps the interactions themselves are better targets than the genes. For example, rather than attempting to reduce *WNT5A* expression globally, a specific targeting of its interaction with *RBMS3* may provide a novel approach to treating SSc-related lung disease without widespread off-target effects. Known genetic interactions may also help predict disease course in individual SSc patients. If it is known, for example, that a patient has the reference *RBMS3* allele and the alternate *WNT5A* allele, this might indicate a highly likelihood of developing severe SSc-related lung disease, and testing and treatment of lung disease could be started early and aggressively. Genetic interactions offer a rich source of information for the understanding and treatment of disease, and new computational methods that take genetic and physiological complexity into account have

the potential to build on the past success of GWAS and to piece together connections between individual findings that are currently isolated from one another.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant R01GM115518. Genotype and phenotype data for the Genome-Wide Association Study in Systemic Sclerosis Study were provided by Maureen D. Mayes, University of Texas Health Science Center, Houston, Texas. Funding support for the original study was provided by the National Institutes of Health, and by other sources detailed in Nat Genet. 2010 May;42(5):426-9. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van 't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, González-Gay MA, González-Escribano MF; Spanish Scleroderma Group, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, Madhok R, Shiels P, Westhovens R, Kreuter A, Kiener H, de Baere E, Witte T, Padykov L, Klareskog L, Beretta L, Scorza R, Lie BA, Hoffmann-Vold AM, Carreira P, Varga J, Hinchcliff M, Gregersen PK, Lee AT, Ying J, Han Y, Weng SF, Amos CI, Wigley FM, Hummers L, Nelson JL, Agarwal SK, Assassi S, Gourh P, Tan FK, Koeleman BP, Arnett FC, Martin J, Mayes MD.

LITERATURE CITED

- Agarwal, S. K., 2010 The genetics of systemic sclerosis. *Discov. Med.* 10: 134–143.
- Agarwal, S. K., P. Gourh, S. Shete, G. Paz, D. Divecha *et al.*, 2009 Association of interleukin 23 receptor polymorphisms with anti-topoisomerase-I positivity and pulmonary hypertension in systemic sclerosis. *J. Rheumatol.* 36: 2715–2723. <https://doi.org/10.3899/jrheum.090421>
- Agarwal, S. K., and J. D. Reveille, 2010 The genetics of scleroderma (systemic sclerosis). *Curr. Opin. Rheumatol.* 22: 133–138. <https://doi.org/10.1097/BOR.0b013e3283367c17>
- Arciniegas, E., A. B. Sutton, T. D. Allen, and A. M. Schor, 1992 Transforming growth factor beta 1 promotes the differentiation of endothelial cells into smooth muscle-like cells in vitro. *J. Cell Sci.* 103: 521–529.
- Assassi, S., T. R. D. J. Radstake, M. D. Mayes, and J. Martin, 2013 Genetics of scleroderma: implications for personalized medicine? *BMC Med.* 11: 9. <https://doi.org/10.1186/1741-7015-11-9>
- Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim *et al.*, 2013 NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41: D991–D995. <https://doi.org/10.1093/nar/gks1193>
- Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Beretta, L., F. Cappiello, J. H. Moore, M. Barili, C. S. Greene *et al.*, 2008a Ability of epistatic interactions of cytokine single-nucleotide polymorphisms to predict susceptibility to disease subsets in systemic sclerosis patients. *Arthritis Rheum.* 59: 974–983. <https://doi.org/10.1002/art.23836>
- Beretta, L., F. Cappiello, J. H. Moore, and R. Scorza, 2008b Interleukin-1 gene complex single nucleotide polymorphisms in systemic sclerosis: a further step ahead. *Hum. Immunol.* 69: 187–192. <https://doi.org/10.1016/j.humimm.2007.12.006>
- Beretta, L., A. Santaniello, M. Mayo, F. Cappiello, M. Marchini *et al.*, 2010 A 3-factor epistatic model predicts digital ulcers in Italian scleroderma patients. *Eur. J. Intern. Med.* 21: 347–353. <https://doi.org/10.1016/j.ejim.2010.05.010>

- Besliu, A., L. Banica, D. Predeteanu, V. Vlad, R. Ionescu *et al.*, 2011 Peripheral blood lymphocytes analysis detects CD100/SEMA4D alteration in systemic sclerosis patients. *Autoimmunity* 44: 427–436. <https://doi.org/10.3109/08916934.2010.541171>
- Betteridge, Z. E., F. Woodhead, H. Lu, G. Shaddick, C. C. Bunn *et al.*, 2016 Brief report: Anti-eukaryotic initiation factor 2b autoantibodies are associated with interstitial lung disease in patients with systemic sclerosis. *Arthritis Rheumatol.* 68: 2778–2783. <https://doi.org/10.1002/art.39755>
- Cappelli, S., S. B. Randone, and G. C. European, 2015 *Interstitial lung disease in systemic sclerosis: where do we stand?* Eur Respiratory Soc, Lausanne, Switzerland.
- Carter, G. W., 2013 Inferring gene function and network organization in *Drosophila* signaling by combined analysis of pleiotropy and epistasis. *G3. Genes, Genomes, and Genetics* 3: 807–814.
- Carter, G. W., M. Hays, A. Sherman, and T. Galitski, 2012 Use of pleiotropy to model genetic interactions in a population. *PLoS Genet.* 8: e1003010. <https://doi.org/10.1371/journal.pgen.1003010>
- Chen, J., D. L.-W. Kwong, C.-L. Zhu, L.-L. Chen, S.-S. Dong *et al.*, 2012 RBMS3 at 3p24 inhibits nasopharyngeal carcinoma development via inhibiting cell proliferation, angiogenesis, and inducing apoptosis. *PLoS One* 7: e44636. <https://doi.org/10.1371/journal.pone.0044636>
- Christmann, R. B., P. Sampaio-Barros, G. Stifano, C. L. Borges, C. R. de Carvalho *et al.*, 2014 Association of Interferon- and transforming growth factor β -regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. *Arthritis Rheumatol.* 66: 714–725. <https://doi.org/10.1002/art.38288>
- Davis, A. P., C. J. Grondin, R. J. Johnson, D. Sciaky, R. McMorran *et al.*, 2019 The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res.* 47: D948–D954. <https://doi.org/10.1093/nar/gky868>
- Davis, A. P., T. C. Wieggers, B. L. King, J. Wieggers, C. J. Grondin *et al.*, 2016 Generating gene ontology-disease inferences to explore mechanisms of human disease at the comparative toxicogenomics database. *PLoS One* 11: e0155530. <https://doi.org/10.1371/journal.pone.0155530>
- Dayem Ullah, A. Z., N. R. Lemoine, and C. Chelala, 2012 SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res.* 40: W65–W70. <https://doi.org/10.1093/nar/gks364>
- Dayem Ullah, A. Z., N. R. Lemoine, and C. Chelala, 2013 A practical guide for the functional annotation of genetic variations using SNPnexus. *Brief. Bioinform.* 14: 437–447. <https://doi.org/10.1093/bib/bbt004>
- Dayem Ullah, A. Z., J. Oscanoa, J. Wang, A. Nagano, N. R. Lemoine *et al.*, 2018 Snpnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.* 46: W109–W113. <https://doi.org/10.1093/nar/gky399>
- Denton, C. P., and D. Khanna, 2017 Systemic sclerosis. *Lancet* 390: 1685–1699. [https://doi.org/10.1016/S0140-6736\(17\)30933-9](https://doi.org/10.1016/S0140-6736(17)30933-9)
- Dissanayake, S. K., M. Wade, C. E. Johnson, M. P. O’Connell, P. D. Leotlela *et al.*, 2007 The Wnt5A/protein kinase c pathway mediates motility in melanoma cells via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal transition. *J. Biol. Chem.* 282: 17259–17271. <https://doi.org/10.1074/jbc.M700075200>
- Distler, O., J. H. Distler, A. Scheid, T. Acker, A. Hirth *et al.*, 2004 Uncontrolled expression of vascular endothelial growth factor and its receptors leads to insufficient skin angiogenesis in patients with systemic sclerosis. *Circ. Res.* 95: 109–116. <https://doi.org/10.1161/01.RES.0000134644.89917.96>
- Edgar, R., M. Domrachev, and A. E. Lash, 2002 Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30: 207–210. <https://doi.org/10.1093/nar/30.1.207>
- Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein, 1998 Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95: 14863–14868. <https://doi.org/10.1073/pnas.95.25.14863>
- Fritz, D., and B. Stefanovic, 2007 RNA-binding protein RBMS3 is expressed in activated hepatic stellate cells and liver fibrosis and increases expression of transcription factor Prx1. *J. Mol. Biol.* 371: 585–595. <https://doi.org/10.1016/j.jmb.2007.06.006>
- Gorlova, O., J.-E. Martin, B. Rueda, B. P. C. Koeleman, J. Ying *et al.*, 2011 Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet.* 7: e1002178. Erratum: 10.1371/annotation/3ae5bb2e-64e5-4548-8d65-1f2d5dfeb073. 10.1371/annotation/7a;52649c-0942-4bd8-a5d3-3cdacca03cd8. <https://doi.org/10.1371/journal.pgen.1002178>
- Hahn, L. W., M. D. Ritchie, and J. H. Moore, 2003 Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 19: 376–382. <https://doi.org/10.1093/bioinformatics/btf869>
- Hashimoto, N., H. Jin, T. Liu, S. W. Chensue, and S. H. Phan, 2004 Bone marrow-derived progenitor cells in pulmonary fibrosis. *J. Clin. Invest.* 113: 243–252. <https://doi.org/10.1172/JCI200418847>
- He, L., X. Zhou, C. Qu, L. Hu, Y. Tang *et al.*, 2014 Musashi2 predicts poor prognosis and invasion in hepatocellular carcinoma by driving epithelial-mesenchymal transition. *J. Cell. Mol. Med.* 18: 49–58. <https://doi.org/10.1111/jcmm.12158>
- Huang, C., D. Liu, J. Nakano, S. Ishikawa, K. Kontani *et al.*, 2005 Wnt5a Expression Is Associated With the Tumor Proliferation and the Stromal Vascular Endothelial Growth Factor—An Expression in Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* 23: 8765–8773. <https://doi.org/10.1200/JCO.2005.02.2871>
- Jimenez, S. A., 2013 Role of endothelial to mesenchymal transition in the pathogenesis of the vascular alterations in systemic sclerosis. *ISRN Rheumatol.* 2013: 835948. <https://doi.org/10.1155/2013/835948>
- Jimenez, S. A., and S. Piera-Velazquez, 2016 Endothelial to mesenchymal transition (EndoMT) in the pathogenesis of systemic sclerosis-associated pulmonary fibrosis and pulmonary arterial hypertension. myth or reality? *Matrix Biol.* 51: 26–36. <https://doi.org/10.1016/j.matbio.2016.01.012>
- Kalluri, R., and R. A. Weinberg, 2009 The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119: 1420–1428. <https://doi.org/10.1172/JCI39104>
- Kano, M., T. Kobayashi, M. Date, M. Tennichi, Y. Hamaguchi *et al.*, 2019 Attenuation of murine sclerodermatous models by the selective SIP 1 receptor modulator cenerimod. *Sci. Rep.* 9: 658. <https://doi.org/10.1038/s41598-018-37074-9>
- Katsumoto, T. R., M. L. Whitfield, and M. K. Connolly, 2011 The pathogenesis of systemic sclerosis. *Annual Review of Pathology: Mechanisms of Disease* 6: 509–537. <https://doi.org/10.1146/annurev-pathol-011110-130312>
- Kim, K. K., M. C. Kugler, P. J. Wolters, L. Robillard, M. G. Galvez *et al.*, 2006 Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. *Proc. Natl. Acad. Sci. USA* 103: 13180–13185. <https://doi.org/10.1073/pnas.0605669103>
- Kudinov, A. E., A. Deneka, A. S. Nikonova, T. N. Beck, Y.-H. Ahn *et al.*, 2016 Musashi-2 (MSI2) supports TGF- β signaling and inhibits claudins to promote non-small cell lung cancer (NSCLC) metastasis. *Proc. Natl. Acad. Sci. USA* 113: 6955–6960. <https://doi.org/10.1073/pnas.1513616113>
- Lafyatis, R., 2012 Connective tissue disease: SSC-fibrosis takes flight with Wingless inhibition. *Nat. Rev. Rheumatol.* 8: 441–442. <https://doi.org/10.1038/nrrheum.2012.99>
- Lam, A. P., A. S. Flozak, S. Russell, J. Wei, M. Jain *et al.*, 2011 Nuclear β -catenin is increased in systemic sclerosis pulmonary fibrosis and promotes lung fibroblast migration and proliferation. *Am. J. Respir. Cell Mol. Biol.* 45: 915–922. <https://doi.org/10.1165/rcmb.2010-0113OC>
- Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559. <https://doi.org/10.1186/1471-2105-9-559>
- Liang, Y., Y. Liu, Q. Meng, X. Li, F. Wang *et al.*, 2015 RBMS3 is a tumor suppressor gene that acts as a favorable prognostic marker in lung

- squamous cell carcinoma. *Med. Oncol.* 32: 459. <https://doi.org/10.1007/s12032-014-0459-9>
- Mahoney, J. M., J. Taroni, V. Martyanov, T. A. Wood, C. S. Greene *et al.*, 2015 Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. *PLOS Comput. Biol.* 11: e1004005. Erratum: e1004133. <https://doi.org/10.1371/journal.pcbi.1004005>
- Manetti, M., S. Guiducci, L. Ibba-Manneschi, and M. Matucci-Cerinic, 2010 Mechanisms in the loss of capillaries in systemic sclerosis: angiogenesis vs. vasculogenesis. *J. Cell. Mol. Med.* 14: 1241–1254. <https://doi.org/10.1111/j.1582-4934.2010.01027.x>
- Manetti, M., E. Romano, I. Rosa, S. Guiducci, S. Bellando-Randone *et al.*, 2017 Endothelial-to-mesenchymal transition contributes to endothelial dysfunction and dermal fibrosis in systemic sclerosis. *Ann. Rheum. Dis.* 76: 924–934. <https://doi.org/10.1136/annrheumdis-2016-210229>
- Marangoni, R. G., B. D. Korman, J. Wei, T. A. Wood, L. V. Graham *et al.*, 2015 Myofibroblasts in murine cutaneous fibrosis originate from adiponectin-positive intradermal progenitors. *Arthritis Rheumatol.* 67: 1062–1073. <https://doi.org/10.1002/art.38990>
- Mayes, M. D., 2012 The genetics of scleroderma: looking into the postgenomic era. *Curr. Opin. Rheumatol.* 24: 677–684. <https://doi.org/10.1097/BOR.0b013e328358575b>
- Nakamura, M., and Y. Tokura, 2011 Epithelial–mesenchymal transition in the skin. *J. Dermatol. Sci.* 61: 7–13. <https://doi.org/10.1016/j.jdermsci.2010.11.015>
- Pattanaik, D., and A. E. Postlethwaite, 2010 A role for lysophosphatidic acid and sphingosine 1-phosphate in the pathogenesis of systemic sclerosis. *Discov. Med.* 10: 161–167.
- Penkov, D., R. Ni, C. Else, S. Piñol-Roma, F. Ramirez *et al.*, 2000 Cloning of a human gene closely related to the genes coding for the c-myc single-strand binding proteins. *Gene* 243: 27–36. [https://doi.org/10.1016/S0378-1119\(99\)00515-6](https://doi.org/10.1016/S0378-1119(99)00515-6)
- Philip, V. M., A. L. Tyler, and G. W. Carter, 2014 Dissection of complex gene expression using the combined analysis of pleiotropy and epistasis. *Pacific Symposium on Biocomputing* pp. 200–211.
- Piera-Velazquez, S., Z. Li, and S. A. Jimenez, 2011 Role of endothelial-mesenchymal transition (EndoMT) in the pathogenesis of fibrotic disorders. *Am. J. Pathol.* 179: 1074–1080. <https://doi.org/10.1016/j.ajpath.2011.06.001>
- Plenge, R. M., M. Seielstad, L. Padyukov, A. T. Lee, E. F. Remmers *et al.*, 2007 TRAF1–C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N. Engl. J. Med.* 357: 1199–1209. <https://doi.org/10.1056/NEJMoa073491>
- Pongracz, J. E., and R. A. Stockley, 2006 Wnt signalling in lung development and diseases. *Respir. Res.* 7: 15. <https://doi.org/10.1186/1465-9921-7-15>
- Postlethwaite, A. E., H. Shigemitsu, and S. Kanangat, 2004 Cellular origins of fibroblasts: possible implications for organ fibrosis in systemic sclerosis. *Curr. Opin. Rheumatol.* 16: 733–738. <https://doi.org/10.1097/01.bor.0000139310.77347.9c>
- Reimand, J., T. Arak, P. Adler, L. Kolberg, S. Reisberg *et al.*, 2016 g: Profiler-A web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 44: W83–W89. <https://doi.org/10.1093/nar/gkw199>
- Rimar, D., Y. Nov, I. Rosner, G. Slobodin, M. Rozenbaum *et al.*, 2015 Semaphorin 3A: an immunoregulator in systemic sclerosis. *Rheumatol. Int.* 35: 1625–1630. <https://doi.org/10.1007/s00296-015-3269-2>
- Schmidt, K. G., M. Herrero San Juan, S. Trautmann, L. Berninger, A. Schwiebs *et al.*, 2017 Sphingosine-1-phosphate receptor 5 modulates early-stage processes during fibrogenesis in a mouse model of systemic sclerosis: a pilot study. *Front. Immunol.* 8: 1242. <https://doi.org/10.3389/fimmu.2017.01242>
- Sherry, S. T., M.-H. Ward, M. Kholodov, J. Baker, L. Phan *et al.*, 2001 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29: 308–311. <https://doi.org/10.1093/nar/29.1.308>
- Solomon, J. J., A. L. Olson, A. Fischer, T. Bull, K. K. Brown *et al.*, 2013 Scleroderma lung disease. *Eur. Respir. Rev.* 22: 6–19. <https://doi.org/10.1183/09059180.00005512>
- Song, L.-W., H.-C. Chen, Y.-F. Lin, J.-H. Yang, C.-C. Chang *et al.*, 2016 Identification of susceptibility gene associated with female primary Sjögren’s syndrome in han chinese by genome-wide association study. *Hum. Genet.* 135: 1287–1294. <https://doi.org/10.1007/s00439-016-1716-0>
- Steen, V. D., and T. A. Medsger, 2007 Changes in causes of death in systemic sclerosis, 1972–2002. *Ann. Rheum. Dis.* 66: 940–944. <https://doi.org/10.1136/ard.2006.066068>
- Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Tokumura, A., L. D. Carbone, Y. Yoshioka, J. Morishige, M. Kikuchi *et al.*, 2009 Elevated serum levels of arachidonoyl-lysophosphatidic acid and sphingosine 1-phosphate in systemic sclerosis. *Int. J. Med. Sci.* 6: 168–176. <https://doi.org/10.7150/ijms.6.168>
- Tsou, P.-S., and A. H. Sawalha, 2017 Unfolding the pathogenesis of scleroderma through genomics and epigenomics. *J. Autoimmun.* 83: 73–94. <https://doi.org/10.1016/j.jaut.2017.05.004>
- Tyler, A. L., L. R. Donahue, G. A. Churchill, and G. W. Carter, 2016 Weak epistasis generally stabilizes phenotypes in a mouse intercross. *PLoS Genet.* 12: e1005805–e1005822. <https://doi.org/10.1371/journal.pgen.1005805>
- Tyler, A. L., B. Ji, D. M. Gatti, S. C. Munger, G. A. Churchill *et al.*, 2017 Epistatic networks jointly influence phenotypes related to metabolic disease and gene expression in diversity outbred mice. *Genetics* 206: 621–639. <https://doi.org/10.1534/genetics.116.198051>
- Tyler, A. L., W. Lu, J. J. Hendrick, V. M. Philip, and G. W. Carter, 2013 CAPE: an R package for combined analysis of pleiotropy and epistasis. *PLOS Comput. Biol.* 9: e1003270. <https://doi.org/10.1371/journal.pcbi.1003270>
- Tyler, A. L., T. C. McGarr, B. J. Beyer, W. N. Frankel, and G. W. Carter, 2014 A genetic interaction network model of a complex neurological disease. *Genes Brain Behav.* 13: 831–840. <https://doi.org/10.1111/gbb.12178>
- Vuga, L. J., A. Ben-Yehudah, E. Kovkarova-Naumovski, T. Oriss, K. F. Gibson *et al.*, 2009 WNT5A is a regulator of fibroblast proliferation and resistance to apoptosis. *Am. J. Respir. Cell Mol. Biol.* 41: 583–589. <https://doi.org/10.1165/rcmb.2008-0201OC>
- Wang, B., Z. Tang, H. Gong, L. Zhu, and X. Liu, 2017 Wnt5a promotes epithelial-to-mesenchymal transition and metastasis in non-small cell lung cancer. *Biosci. Rep.* 37: 1089–1100. <https://doi.org/10.1042/BSR20171092>
- Wang, M.-H., S.-Y. Qin, S.-G. Zhang, G.-X. Li, Z.-H. Yu *et al.*, 2015 Musashi-2 promotes hepatitis B virus related hepatocellular carcinoma progression via the Wnt/ β -catenin pathway. *Am. J. Cancer Res.* 5: 1089.
- Wang, W., X. Yu, C. Wu, and H. Jin, 2018 Differential effects of Wnt5a on the proliferation, differentiation and inflammatory response of keratinocytes. *Mol. Med. Rep.* 17: 4043–4048.
- Wei, J., D. Melichian, K. Komura, M. Hinchcliff, A. P. Lam *et al.*, 2011 Canonical Wnt signaling induces skin fibrosis and subcutaneous lipotrophy: a novel mouse model for scleroderma? *Arthritis Rheum.* 63: 1707–1717. <https://doi.org/10.1002/art.30312>
- Whitfield, M. L., G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball *et al.*, 2002 Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* 13: 1977–2000. <https://doi.org/10.1091/mbc.02-02-0030>
- Willis, B. C., and Z. Borok, 2007 TGF- β -induced EMT: mechanisms and implications for fibrotic lung disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 293: L525–L534. <https://doi.org/10.1152/ajplung.00163.2007>
- Yang, Y., L. Quan, and Y. Ling, 2018 RBMS3 inhibits the proliferation and metastasis of breast cancer cells. *Oncology Research Featuring Preclinical*

- and *Clinical Cancer Therapeutics* 26: 9–15. <https://doi.org/10.3727/096504017X14871200709504>
- Yu, W., M. Clyne, M. J. Khoury, and M. Gwinn, 2010 Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* 26: 145–146. <https://doi.org/10.1093/bioinformatics/btp618>
- Zeisberg, E. M., O. Tarnavski, M. Zeisberg, A. L. Dorfman, J. R. McMullen *et al.*, 2007 Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat. Med.* 13: 952–961. <https://doi.org/10.1038/nm1613>
- Zhu, Q., A. K. Wong, A. Krishnan, M. R. Aure, A. Tadych *et al.*, 2015 Targeted exploration and analysis of large cross-platform human transcriptomic compendia. *Nat. Methods* 12: 211–214. <https://doi.org/10.1038/nmeth.3249>
- Zhu, S., and G. Fang, 2018 MatrixEpistasis: ultrafast, exhaustive epistasis scan for quantitative traits with covariate adjustment. *Bioinformatics* 34: 2341–2348. <https://doi.org/10.1093/bioinformatics/bty094>

Communicating editor: C. Myers