ARTHRITIS & RHEUMATOLOGY
Vol. 70, No. 10, October 2018, pp 1654–1660
DOI 10.1002/art.40541

BRIEF REPORT

Whole-Exome Sequencing to Identify Rare Variants and Gene Networks That Increase Susceptibility to Scleroderma in African Americans

Pravitt Gourh,[1] Elaine F. Remmers,[2] Steven E. Boyden,[2] Theresa Alexander,[3] Nadia D. Morgan,[4] Ami A. Shah,[4] Maureen D. Mayes,[5] Ayo Doumatey,[2] Amy R. Bentley,[2] Daniel Shriner,[2] Robyn T. Domsic,[6] Thomas A. Medsger Jr.,[6] Virginia D. Steen,[7] Paula S. Ramos,[8] Richard M. Silver,[8] Benjamin Korman,[9] John Varga,[9] Elena Schiopu,[10] Dinesh Khanna,[10] Vivien Hsu,[11] Jessica K. Gordon,[12] Lesley Ann Saketkoo,[13] Heather Gladue,[14] Brynn Kron,[15] Lindsey A. Criswell,[15] Chris T. Derk,[16] S. Louis Bridges Jr.,[17] Victoria K. Shanmugam,[18] Kathleen D. Kolstad,[19] Lorinda Chung,[20] Reem Jan,[21] Elana J. Bernstein,[22] Avram Goldberg,[23] Marcin Trojanowski,[24] Suzanne Kafaja,[25] Kathleen M. Maksimowicz-McKinnon,[26] James C. Mullikin,[27] Adebowale Adeyemo,[2] Charles Rotimi,[2] Francesco Boin,[15] Daniel L. Kastner,[2] and Fredrick M. Wigley[4]

Objective. **Whole-exome sequencing (WES) studies in systemic sclerosis (SSc) patients of European American (EA) ancestry have identified variants in the *ATP8B4* gene and enrichment of variants in genes in the extracellular matrix (ECM)–related pathway that increase SSc susceptibility. This study was undertaken to evaluate the association of the *ATP8B4* gene and the ECM-related pathway with SSc in a cohort of African American (AA) patients.**

Methods. **SSc patients of AA ancestry were enrolled from 23 academic centers across the US under the Genome Research in African American Scleroderma Patients consortium. Unrelated AA individuals without serologic evidence of autoimmunity who were enrolled in the Howard University Family Study were used as unaffected controls. Functional variants in genes reported in the 2 WES studies in EA patients with SSc were selected for gene association**

Dr. Gourh is recipient of a Scientist Development award from the Rheumatology Research Foundation. Ms Morgan's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases [NIAMS] grant T32-AR-048522) and the Rheumatology Research Foundation (Scientist Development award). Dr. Shah's work was supported by the NIH (NIAMS grant K23-AR-061439). Dr. Mayes' work was supported by NIAMS grant P50-AR-054144 from the Centers of Research Translation, NIH grants N01-AR-02251 and R01-AR-055258, and the Department of Defense Congressionally Directed Medical Research Program (grants W81XWH-07–1–011 and WX81XWH-13–1-0452). Dr. Ramos' work was supported by the NIH (grants K01-AR-067280, R03-AR-065801, and P60-AR-062755) and the South Carolina Clinical and Translational Research Institute (Medical University of South Carolina) through the NIH (grants UL1-RR-029882 and UL1-TR-000062). Dr. Silver's work was supported by the NIH (grant P60-AR-062755) and the South Carolina Clinical and Translational Research Institute (Medical University of South Carolina) through the NIH (grants UL1-RR-029882 and UL1-TR-000062). Dr. Khanna's work was supported by NIAMS grant K24-AR-063120. Dr. Shanmugam's work was supported by the NIH (grant R01-NR-013888 from the National Institute of Nursing Research), the Defense Advanced Research Projects Agency, and the Scleroderma Research Foundation. Dr. Boin's work was supported by the Nina Ireland Program for Lung Health.

[1]Pravitt Gourh, MD: NIAMS and National Human Genome Research Institute, NIH, Bethesda, Maryland; [2]Elaine F. Remmers, PhD, Steven E. Boyden, PhD, Ayo Doumatey, PhD, Amy R. Bentley, PhD, Daniel Shriner, PhD, Adebowale Adeyemo, MD, Charles Rotimi, PhD, Daniel L. Kastner, MD, PhD: National Human Genome Research Institute, NIH, Bethesda, Maryland; [3]Theresa Alexander, MS: NIAMS, NIH, Bethesda, Maryland; [4]Nadia D. Morgan, MBBS, MHS, Ami A. Shah, MD, MHS, Fredrick M. Wigley, MD: Johns Hopkins University School of Medicine, Baltimore, Maryland; [5]Maureen D. Mayes, MD, PhD: University of Texas McGovern Medical School, Houston; [6]Robyn T. Domsic, MD, MPH, Thomas A. Medsger Jr., MD: University of Pittsburgh, Pittsburgh, Pennsylvania; [7]Virginia D. Steen, MD: Georgetown University School of Medicine, Washington, DC; [8]Paula S. Ramos, PhD, Richard M. Silver, MD: Medical University of South Carolina, Charleston; [9]Benjamin Korman, MD, John Varga, MD: Northwestern University Feinberg School of Medicine, Chicago, Illinois; [10]Elena Schiopu, MD, Dinesh Khanna, MD, MSc: University of Michigan, Ann Arbor; [11]Vivien Hsu, MD: Robert Wood Johnson University, New Brunswick, New Jersey; [12]Jessica K. Gordon, MD, MSc: Hospital for Special Surgery, New York, New York; [13]Lesley Ann Saketkoo, MD, MPH: Tulane University School of Medicine, New Orleans, Louisiana; [14]Heather Gladue, DO: Arthritis and Osteoporosis Consultants of the Carolinas, Charlotte, North Carolina; [15]Brynn Kron, BS, Lindsey A. Criswell, MD, MPH, Francesco Boin, MD: University of California, San Francisco; [16]Chris T. Derk, MD, MS: University of Pennsylvania, Philadelphia; [17]S. Louis Bridges Jr., MD, PhD: University of Alabama at Birmingham; [18]Victoria K. Shanmugam, MD, MRCP: The George Washington University, Washington, DC; [19]Kathleen D. Kolstad, MD, PhD: Stanford University School of Medicine, Stanford, California; [20]Lorinda Chung, MD, MS: Stanford University School of Medicine, Stanford, California, and Palo Alto VA Health Care System, Palo Alto, California; [21]Reem Jan, MBBS: University of Chicago Pritzker School of Medicine, Chicago, Illinois; [22]Elana J. Bernstein, MD, MS: New York Presbyterian Hospital, Columbia University, New York, New York; [23]Avram Goldberg, MD: New York University Langone Medical Center, New York, New York; [24]Marcin Trojanowski, MD: Boston University School of Medicine,

testing using the optimized sequence kernel association test (SKAT-O) and pathway analysis by Ingenuity Pathway Analysis in 379 patients and 411 controls.

***Results*. Principal components analysis demonstrated that the patients and controls had similar ancestral backgrounds, with roughly equal proportions of mean European admixture. Using SKAT-O, we examined the association of individual genes that were previously reported in EA patients and none remained significant, including *ATP8B4* ($P = 0.98$). However, we confirmed the previously reported association of the ECM-related pathway with enrichment of variants within the *COL13A1*, *COL18A1*, *COL22A1*, *COL4A3*, *COL4A4*, *COL5A2*, *PROK1*, and *SERPINE1* genes (corrected $P = 1.95 \times 10^{-4}$).**

***Conclusion*. In the largest genetic study in AA patients with SSc to date, our findings corroborate the role of functional variants that aggregate in a fibrotic pathway and increase SSc susceptibility.**

Systemic sclerosis (SSc; scleroderma) is a chronic multisystem disease, characterized by fibrosis of the skin and internal organs, systemic vasculopathy, and autoimmunity. Compared with individuals of European American (EA) ancestry, African Americans (AA) in the US have a higher incidence and prevalence of SSc (1). SSc occurs at an earlier age in persons of AA descent compared to EA and is more likely to be manifested as diffuse skin involvement and the presence of anti–topoisomerase I or antifibrillarin antibodies—features that are associated with severe disease and a worse outcome (2). AA individuals are more likely to develop severe interstitial lung disease (ILD) or pulmonary arterial hypertension. This lung disease accounts for the major overall SSc-related deaths in all racial groups (3).

The etiology of SSc is unknown, but several environmental agents and genetic variants have been implicated. A strong role of genetic etiologic factors has been suggested by family studies in SSc, which have demonstrated an absolute risk of 1.6% in families as compared to 0.026% in the general population (4). Candidate gene, family-based, and genome-wide association studies (GWAS) investigating common variants, that have been conducted mostly in EA patients with SSc, have revealed autoimmune disease susceptibility loci that are not unique to SSc (5). These studies focused on common variants that have been able to only partially account for SSc heritability. Rare variants (minor allele frequency [MAF] <0.5%) and low-frequency variants (MAF 0.5–5%) have recently been implicated in several diseases and could account for a portion of the missing data on SSc heritability.

Recent reports suggest an increased burden of rare coding variants in genes and pathways of complex diseases beyond the common variants identified by GWAS (6,7). Various techniques used to identify genes with aggregation of deleterious variants include candidate gene Sanger sequencing for a hypothesis-driven approach. Next generation sequencing platforms, including whole-exome sequencing (WES) and whole-genome sequencing, provide hypothesis-neutral approaches. Gao et al (6) recently performed WES in 78 EA patients with SSc and found an enrichment in functional *ATP8B4* variants as compared to controls ($P = 2.77 \times 10^{-7}$). Furthermore, a single missense variant (rs55687265) was associated with SSc ($P = 9.35 \times 10^{-10}$), and when this variant was removed, the *ATP8B4* gene–based association was eliminated. Mak et al (7) performed WES in 32 EA patients with SSc, identifying 70 genes enriched for deleterious variants in the diffuse SSc subset; they reported significant enrichment of variants in the *COL4A3*, *COL4A4*, *COL5A2*, *COL13A1,* and *COL22A1* genes in the extracellular matrix (ECM)–related pathway ($P = 0.002$). Given the potential importance of the findings of these prior smaller WES studies in EA patients with SSc, we investigated their significance in a larger cohort of AA patients with SSc by performing WES gene and pathway-based association testing.

Our understanding of genetic susceptibility in AA patients with SSc is limited and restricted to HLA antigens, due to the lack of extensive studies in the AA population (8). The Genome Research in African American Scleroderma Patients (GRASP) consortium was created

Boston, Massachusetts; [25]Suzanne Kafaja, MD: David Geffen School of Medicine, University of California, Los Angeles; [26]Kathleen M. Maksimowicz-McKinnon, DO: Henry Ford Health System, Detroit, Michigan; [27]James C. Mullikin, PhD: National Human Genome Research Institute, NIH Intramural Sequencing Center, Rockville, Maryland.

Drs. Boin, Kastner, and Wigley contributed equally to this work.

Dr. Mayes has received consulting fees from Mitsubishi Tanabe, Astellas, Boehringer Ingelheim, Gerson Lehrman Group, Smart Analyst, and Guidepoint Global (less than $10,000 each). Dr. Domsic has received consulting fees from Eicos Sciences (less than $10,000). Dr. Khanna has received consulting fees from Actelion, Bristol-Myers Squibb, CSL Behring, Inventiva, EMD Merck-Serono, Sanofi-Aventis, GlaxoSmithKline, Corbus, Cytori, and UCB (less than $10,000 each), Bayer, Boehringer Ingelheim, Corbus, and Genentech-Roche (more than $10,000 each), and research support from Bayer, Bristol-Myers Squibb, and Pfizer, and owns stock or stock options in Eicos Sciences, Inc. (now CiViBioPharma, Inc.). Dr. Saketkoo has received consulting fees from Axon Pharma (less than $10,000). Dr. Gladue has received consulting fees from Pfizer, AbbVie, and Actelion (less than $10,000 each) and from Horizon Pharmaceuticals (more than $10,000). Dr. Derk has received research support from Gilead, Actelion, and Cytori. Dr. Shanmugam has received research support from AbbVie. Dr. Bernstein has received consulting fees from Genentech (less than $10,000).

Address correspondence to Pravitt Gourh, MD, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Human Genome Research Institute, NIH, Building 10, Room 5-SEN-3481, 10 Center Drive, Bethesda, MD 20892. E-mail: pravitt.gourh@nih.gov.

Submitted for publication March 6, 2018; accepted in revised form April 26, 2018.

to assemble a large cohort of AA patients with SSc in order to conduct systematic and comprehensive genetic studies. Within the cohort, 400 AA patients with SSc and 482 controls have undergone WES, and studies of select genes from the WES analysis are being replicated in an independent series for confirmation. Herein we present the results of our analyses for gene-level associations and pathway associations in genes that were previously reported in WES studies of 76 EA patients with SSc by Gao et al and 32 EA patients with SSc by Mak et al (6,7). We replicated a collective enrichment of coding and deleterious variants in genes of the fibrotic pathway in AA patients with SSc.

PATIENTS AND METHODS

Study population. AA patients with SSc were enrolled in the GRASP consortium from 23 academic centers in the US (see Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40541/abstract). Enrolled patients self-identified as African American. All patients met the 1980 American College of Rheumatology (ACR) or the 2013 ACR/European League Against Rheumatism classification criteria for SSc (9,10) or had at least 3 of the 5 features of CREST syndrome (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasias) (11). Control samples were obtained from the Howard University Family Study, a population-based study of AA families and unrelated individuals enrolled in the Washington, DC metropolitan area (12). Only unrelated individuals were included as controls. Sera obtained from controls were tested for antinuclear antibodies (ANAs) by indirect immunofluorescence, and only those with a titer of <1:80 were included in this study. DNA was extracted from samples of whole blood or saliva.

Sequence analysis. WES was performed on 400 SSc and 482 control samples using SeqCap EZ Human Exome + UTR Probes (Roche), and libraries were sequenced on the HiSeq 2000 platform (Illumina) using 2 × 100-bp paired-end reads. WES data were analyzed using the computational resources of the NIH high-performance computing Biowulf cluster (http://hpc.nih.gov). Multiple variant and sample quality control filters were used prior to analyses (see Supplementary Methods, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40541). The sequence data shown in this study are accessible from the Sequence Read Archive under the accession number SRP140756.

Identity by descent, admixture, and principal components analyses. Identity by descent analysis was performed using common variants (MAF >5%) from the sequence data after linkage disequilibrium (LD) pruning (r^2 <0.5) for estimating kinship coefficients. In order to remove familial relatedness, only 1 sample with the highest call rate was included from individuals with pi-hat >0.085. We used the ADMIXTURE software tool (13) and the 1000 Genomes populations as references for estimation of population admixture in our patients and controls. Principal components analysis (PCA) was used to estimate population stratification. A set of 35,280 markers (LD pruned [r^2 <0.5] and MAF >5%) was used to compute principal components and the top ten principal component eigenvalues were used to correct for population stratification using SNP & Variation Suite v8.7.1 (Golden Helix). Samples that clustered within the European cluster were removed from the analysis.

Statistical analysis. Gene-level testing was performed using an optimized sequence kernel association test (SKAT-O) with Madsen and Browning marker weighting and was corrected for ancestry using the top 10 principal components. Functional variants (missense, nonsense, and splice site) with a MAF of <0.05 present on 20 candidate genes as reported by Gao et al (6), and functional variants of all frequencies present on 87 genes from the diffuse disease and ILD subsets of SSc patients analyzed by Mak et al (7) were examined for association in the GRASP cohort (see Supplementary Figure 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40541/abstract). In order to account for multiple gene testing, Bonferroni correction was applied for the total number of tests performed. This included 20 tests for the Gao et al study and 87 for the Mak et al study, which yielded *P* value significance thresholds of 0.0025 and 0.00057, respectively (6,7). Ingenuity Pathway Analysis (IPA) software was used to identify canonical pathways with all of the reported genes in the studies by Gao and colleagues and Mak and colleagues. *P* values for pathways were generated using Fisher's right-tailed exact test and were corrected for multiple testing using the Benjamini-Hochberg false discovery rate criterion.

RESULTS

Patient characteristics. A few controls (2.17%) were ANA+ (titer ≥1:80) and were excluded from the study. As we had hypothesized, there were more female subjects in the SSc cohort as compared to the control cohort (Table 1). Detailed sociodemographic, clinical, and serologic characteristics of the GRASP cohort have been reported by Morgan et al (11). In order to address population stratification, we performed PCA and plots showed that the patients and controls in the GRASP cohort were well-matched and that there was no major stratification at a global genomic level (Figure 1 and Supplementary

Table 1. Clinical and serologic characteristics of the African American patients with SSc and healthy controls*

	SSc patients (n = 379)	Healthy controls (n = 411)
Sex, male/female	64 (16.9)/315 (83.1)	217 (52.8)/194 (47.2)
SSc skin involvement		
Limited	166 (43.8)	–
Diffuse	188 (49.6)	–
Autoantibodies		
Anticentromere	44 (11.6)	–
Anti–topoisomerase I	107 (28.2)	–
Pulmonary arterial hypertension	73 (19.3)†	–
Interstitial lung disease	114 (30.1)‡	–

* Values are the number (%) of patients with systemic sclerosis (SSc) and unaffected controls.
† Data are based on right-sided heart catheterization.
‡ Data are based on computed tomography of the chest.

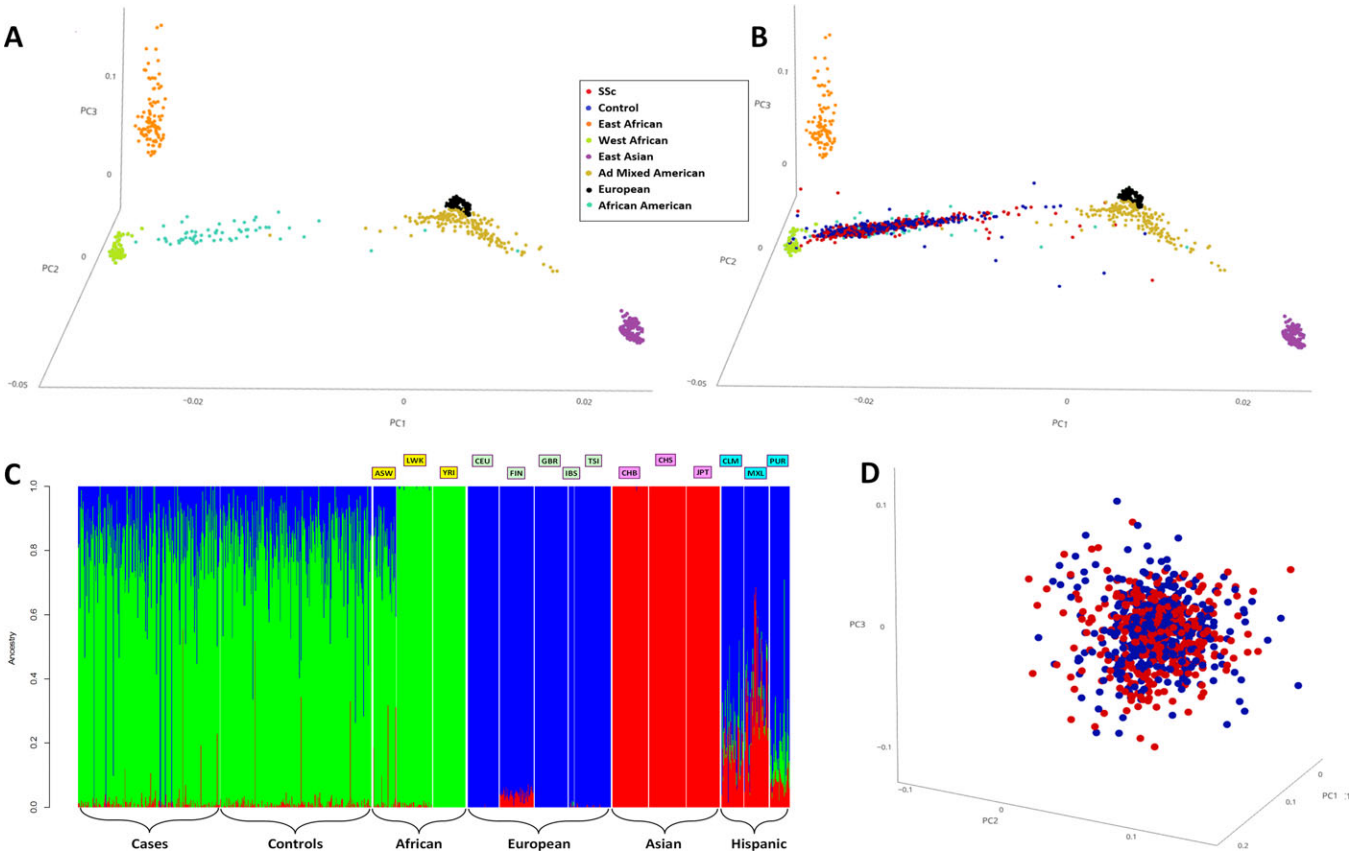

Figure 1. **A** and **B,** Three-dimensional principal components analysis (PCA) plots of the 1000 Genomes populations (**A**) and the 1000 Genomes populations along with the systemic sclerosis (SSc) patients and controls from the present study (**B**). Each dot represents an individual subject. Patients with SSc (red), controls (blue), Americans of admixture (Colombians from Medellin, Colombia [CLM], subjects of Mexican Ancestry from Los Angeles, California [MXL], and Puerto Ricans from Puerto Rico [PUR]) (gold), African Americans (light blue), East Africans (Luhya in Webuye, Kenya [LWK]) (orange), East Asians (Han Chinese in Beijing [CHB], Southern Han Chinese [CHS], and Japanese in Tokyo, Japan [JPT]) (purple), Europeans (Utah Residents of North and Western Europe ancestry [CEU], Finnish in Finland [FIN], British in England and Scotland [GBR], Iberian population in Spain [IBS], and Toscani in Italy [TSI]) (black), and West Africans (Yoruba in Ibadan, Nigeria [YRI]) (light green) are shown. **C**, Admixture plot of the 1000 Genomes populations along with the SSc patients and controls from the present study. Each individual subject is represented as a vertical bar. The y-axis depicts contributions from African (green), European (blue), Asian (red), and Hispanic ancestries. The results in cases and controls are similar to each other and similar to those in the Americans of African Ancestry in southwestern US (ASW) group from the 1000 Genomes project. **D**, Three-dimensional PCA with the top 3 principal components. Each dot represents an individual subject. Patients with SSc (red) and controls (blue) are shown.

Figures 2 and 3, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract).

Results of sequence analysis. A total of 400 SSc and 482 control exomes were sequenced using the same platform and were analyzed simultaneously. After filtration for quality control, 379 SSc cases and 411 controls remained and were used for further analysis (see Supplementary Methods, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). An average of 90% of targeted bases produced high-confidence calls, and mean depth of coverage was 47× in the targeted region. The transition-to-transversion ratio for the coding region was 3.29, and the ratio for heterozygous-to-nonreference homozygous variants was 2.6.

PCA was performed for fine characterization of genetic ancestry. A 3-dimensional PCA plot depicted the AA patients to be spread between the West African and European clusters based on their degree of European admixture (Figures 1A and B, and Supplementary Figures 2 and 3, http://online.library.wiley.com/doi/10.1002/art.40541/abstract). The GRASP samples were closer to the West African cluster and distinct from the East African cluster, confirming West Africa as the main ancestral home of the AA patients. The admixture proportions in GRASP samples were similar to those in AA persons in the 1000 Genomes Project (Americans of African Ancestry in southwestern US), with major contributions from West African and European ancestries and a minor

contribution from Asian ancestry, likely representing Native American admixture (Figure 1C). The mean ± SD individual European admixture was 16.98 ± 12.4% in the SSc samples and 16.52 ± 11.7% in controls, and the difference was not statistically significant ($P = 0.59$, by *t*-test) (see Supplementary Figure 4, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). The PCA plot also confirmed that the patients and controls were very ancestrally similar to each other (Figure 1D).

Gene-level analysis. All variants in the 20 genes, as described in the report by Gao and colleagues (6), were included for gene-level testing by SKAT-O with the top 10 principal components as covariates. The *ATP8B4* gene-level association reported in EA patients did not reach statistical significance in the GRASP cohort ($P = 0.98$) (see Supplementary Tables 2 and 3, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). A missense variant (rs55687265), which was found to be the primary signal for association in the *ATP8B4* gene in the study by Gao et al (6), had similar frequency in the SSc patients and controls and was not statistically significant in the GRASP cohort ($P = 0.84$) (Supplementary Table 4, http://online.library.wiley.com/doi/10.1002/art.40541/abstract).

The 87 unique genes that were identified in the study by Mak and colleagues (7) were evaluated in the GRASP cohort by SKAT-O. For the *COL4A4* gene, the *P* value was 0.038, but this did not remain significant after correction for multiple testing (see Supplementary Table 5, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). Upon examining the subset of SSc patients with diffuse disease, we discovered the *MSR1* gene ($P = 0.01$), and when evaluating the subset of SSc patients with ILD we identified the *ZNF492* gene ($P = 0.01$), the *FOLR3* gene ($P = 0.03$), and the *STAB1* gene ($P = 0.04$). After correction for multiple testing, however, these associations were not statistically significant (see Supplementary Tables 6 and 7, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). The growth differentiation factor 2 gene (*GDF-2*) was reported by Mak et al to have a potential enrichment of variants by burden ratio (7) and was significant with an uncorrected *P* value of 0.00029 in the study by Gao et al (6). In the GRASP cohort this gene initially showed significance (uncorrected $P = 0.04$), but it did not remain significant after correction for multiple testing (see Supplementary Table 2, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract).

Pathway-based analysis. IPA was used to predict pathways enriched with gene variants. Gene-level association results from the SKAT-O analysis of the 20 genes from the study by Gao et al and the 87 genes from the study by Mak et al (6,7), in all SSc patients, were used for pathway prediction. None of the pathways that were predicted based on the gene list from the study by Gao and colleagues were statistically significant after correction for multiple testing. The hepatic fibrosis/hepatic stellate cell activation pathway, identified with IPA based on the gene list in the study by Mak and colleagues, comprised not only the *COL13A1, COL22A1, COL4A3, COL4A4,* and *COL5A2* genes, but additionally included the *COL18A1, PROK1,* and *SERPINE1* genes belonging to the same pathway (see Supplementary Table 8, available at http://online.library.wiley.com/doi/10.1002/art.40541/abstract). In the GRASP cohort, this was the only pathway that was statistically significant after correction for multiple testing (corrected $P = 1.95 \times 10^{-4}$) (Table 2). After correction for multiple testing, the hepatic fibrosis/hepatic stellate cell activation pathway was the only significant pathway in the subset of SSc patients with diffuse disease (Table 2).

Table 2. Pathway analysis of the candidate genes in the GRASP cohort*

Ingenuity pathway	*P*†	Corrected *P*‡
All SSc patients versus controls		
Hepatic fibrosis/hepatic stellate cell activation	2.09×10^{-6}	1.95×10^{-4}
Melatonin degradation III	0.005	0.22
Coagulation system	0.01	0.29
Complement system	0.01	0.29
Atherosclerosis signaling	0.02	0.34
Diffuse SSc subset of patients versus controls		
Hepatic fibrosis/hepatic stellate cell activation	2.27×10^{-6}	2.14×10^{-4}
Melatonin degradation III	0.005	0.22
Coagulation system	0.01	0.29
Complement system	0.01	0.29
Atherosclerosis signaling	0.02	0.36

* Data depict the top 5 pathways and results from the optimized sequence kernel association test of the 87 genes (from the study by Mak et al [7]) in all patients with systemic sclerosis (SSc) and in the diffuse SSc patient subset. The Ingenuity Pathway Analysis program was used for predicting pathways. GRASP = Genome Research in African American Scleroderma Patients.
† By Fisher's right-tailed exact test.
‡ Corrected for multiple testing using the Benjamini-Hochberg correction.

DISCUSSION

In this large cohort of AA patients with SSc, we examined the 20 candidate genes as identified by Gao et al and the 87 candidate genes as identified by Mak et al in their recent WES studies of EA patients with SSc (6,7). Despite having a much larger sample size than these original studies, we were unable to replicate the *ATP8B4* gene association or the rs55687265 variant in the *ATP8B4* gene in the present study. We were, however, able to replicate the association with the ECM-related pathway, but found no additional associations.

One of the greatest accomplishments achieved as a result of this investigation is the establishment of the GRASP cohort, which comprised a large sample of AA patients with SSc and controls with similar ancestral backgrounds. The GRASP cohort will serve as a valuable resource for future transethnic genomic studies in SSc.

Based on the genes reported as candidates for association in the study by Mak et al (7), pathway analysis in the GRASP cohort highlighted a fibrotic pathway that had enrichment of genes with functional variants involved in ECM biology. The gene list included several genes in the collagen family and genes involved in fibrinolysis and angiogenesis. The clustering of these genes into a fibrosis network corresponds to the excessive synthesis and deposition of ECM proteins observed in SSc and, thus, could be a potential candidate for targeted therapy. The *GDF2* gene (also known as bone morphogenetic protein 9) that was evaluated in the present study and identified in the studies by Gao et al and Mak et al (6,7) did not reach statistical significance after correction for multiple testing, but remains an interesting candidate gene involved in modulation of the ECM.

In the GRASP cohort, the ancestries of AA patients as well as the controls were primarily derived from West Africa, and on average the AA patients and controls had similar proportions of African, European, and Asian (likely Native American) descent. The patients and controls were recruited from different geographic locations in the US and, using PCA and the ADMIXTURE software tool, we were able to demonstrate that there was no major population stratification between the patients and controls.

Ancestry-specific associations have previously been identified in complex diseases. Similar to the *PADI4* gene association in Asians and the *PTPN22* gene association in Europeans with rheumatoid arthritis, it may be the case that there are differences in SSc susceptibility loci in different ancestral populations (14). The admixture of the AA population is a recent event in human admixture history and, as we have demonstrated, contains varying amounts of European and Asian ancestries. This difference in genetic architecture could explain the lack of association of the *ATP8B4* gene and the rs55687265 variant in AA patients with SSc.

In the original report of the association of the rs55687265 variant with scleroderma in EA patients by Gao and colleagues (6), the discovery set contained 78 patients in whom the likelihood of association was significant (odds ratio [OR] 6.11, $P = 9.35 \times 10^{-10}$) and the replication set had 415 patients in whom a significant association was found (OR 1.86, $P = 0.01$). Nevertheless, a recent study of 7,426 SSc patients and 13,087 controls of European ancestry was also unable to replicate the association with the rs55687265 variant (15). This underscores the importance of replication in genetics studies in large well-established cohorts to demonstrate reproducibility, provide a better estimate of effect size, and confirm that the original association is not due to unidentified biases present in a single study. We expect that, upon completion of the GRASP targeted resequencing and final analysis of ~400 genes, at least a few of these genes will show a statistically significant association with scleroderma and increase our understanding of molecular pathways involved in SSc pathogenesis.

ACKNOWLEDGMENT

We thank Dr. Kalpana Manthiram for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Gourh had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Gourh, Remmers, Boin, Kastner, Wigley.

Acquisition of data. Gourh, Remmers, Morgan, Shah, Mayes, Doumatey, Bentley, Shriner, Domsic, Medsger, Steen, Ramos, Silver, Korman, Varga, Schiopu, Khanna, Hsu, Gordon, Saketkoo, Gladue, Kron, Criswell, Derk, Bridges, Shanmugam, Kolstad, Chung, Jan, Bernstein, Goldberg, Trojanowski, Kafaja, Maksimowicz-McKinnon, Mullikin, Adeyemo, Rotimi, Boin, Kastner, Wigley.

Analysis and interpretation of data. Gourh, Remmers, Boyden, Alexander.

REFERENCES

1. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48:2246–55.
2. Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB, Medsger TA Jr, Lucas M, Michet CJ, et al. Mortality in systemic sclerosis: an international meta-analysis of individual patient data. Am J Med 2005;118:2–10.
3. Poudel DR, Jayakumar D, Danve A, Sehra ST, Derk CT. Determinants of mortality in systemic sclerosis: a focused review. Rheumatol Int 2017. E-pub ahead of print.
4. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis Rheum 2001;44:1359–62.
5. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010;42:426–9.
6. Gao L, Emond MJ, Louie T, Cheadle C, Berger AE, Rafaels N, et al. Identification of rare variants in ATP8B4 as a risk factor for systemic sclerosis by whole-exome sequencing. Arthritis Rheumatol 2016;68:191–200.
7. Mak AC, Tang PL, Cleveland C, Smith MH, Connolly MK, Katsumoto TR, et al. Whole-exome sequencing for identification of potential causal variants for diffuse cutaneous systemic sclerosis. Arthritis Rheumatol 2016;68:2257–62.

8. Arnett FC, Gourh P, Shete S, Ahn CW, Honey RE, Agarwal SK, et al. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. Ann Rheum Dis 2010;69:822–7.
9. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
10. Van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.
11. Morgan ND, Shah AA, Mayes MD, Domsic RT, Medsger TA Jr, Steen VD, et al. Clinical and serological features of systemic sclerosis in a multicenter African American cohort: analysis of the genome research in African American scleroderma patients clinical database. Medicine (Baltimore) 2017;96:e8980.
12. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, et al. A genome-wide association study of hypertension and blood pressure in African Americans. PLoS Genet 2009;5:e1000564.
13. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–64.
14. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. Nat Rev Rheumatol 2010;6:290–5.
15. López-Isac E, Bossini-Castillo L, Palma AB, Assassi S, Mayes MD, Simeón CP, et al. Analysis of ATP8B4 F436L missense variant in a large systemic sclerosis cohort. Arthritis Rheumatol 2017;69:1337–8.