Associations of Multiple *NOTCH4* Exonic Variants with Systemic Sclerosis

Xiaodong Zhou, Hongye Li, Shicheng Guo, Jiucun Wang, Chunhua Shi, Maribel Espitia, Xinjian Guo, Qingwen Wang, Mengyuan Liu, Shervin Assassi, John D. Reveille, and Maureen D. Mayes

ABSTRACT. ***Objective.*** Findings from previous genome-wide association studies indicated an association of the *NOTCH4* gene with systemic sclerosis (SSc). This is a followup study to fine-map exonic variants of *NOTCH4* in SSc.

Methods. All exons of *NOTCH4* were sequenced and analyzed in a total of 1006 patients with SSc and 1004 controls of US white ancestry with the Ion Torrent system. Identified SSc-associated variants were confirmed with Sanger sequencing, and then examined in a Chinese Han cohort consisting of 576 patients with SSc and 574 controls. The *NOTCH4* variants were analyzed for association with SSc as a whole and with SSc clinical and autoantibody subtypes with and without the influence of specific HLA-class II alleles that had been previously identified as major genetic factors in SSc.

Results. A total of 12 SSc-associated and SSc subtype–associated exonic variants of *NOTCH4* were identified in the US cohort. Three of them are nonsynonymous single-nucleotide polymorphisms and 1 is a CTG tandem repeat that encodes for a poly-leucine, all of which are located in the *NOTCH4* extracellular domain (NECD). Conditional logistic regression analysis on SSc-associated HLA-class II alleles indicated an independent association of the *NOTCH4* variants with SSc autoantibody subtypes. Analysis of the Chinese cohort supported a genetic contribution of *NOTCH4* to SSc and its subtypes.

Conclusion. Multiple *NOTCH4* exonic variants were associated with SSc and/or SSc subtypes. Several of these variants encode nonsynonymous sequence changes occurring in the NECD, which implicates a potentially functional effect of *NOTCH4*. (J Rheumatol First Release November 15 2018; doi:10.3899/jrheum.180094)

Key Indexing Terms:
SYSTEMIC SCLEROSIS | *NOTCH4* GENE | EXONIC VARIANTS

From the Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; Human Genetics Center, Division of Biostatistics, The University of Texas School of Public Health, Houston, Texas, USA; State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai; Department of Rheumatology, Jiangxi People's Hospital, Nanchang; Department of Rheumatology, Peking University-Shenzhen Hospital, Shenzhen; Life Sciences College, Hubei University, Wuhan, China.

Supported by the US National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases 1U01AI09090-01 and NIH P01-052915-01; the Major National Science and Technology Program of China, grant number 2008ZX10002-002; and the Science and Technology Committee of Shanghai Municipality (11410701800).

X. Zhou, MD, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; H. Li, MD, PhD, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; S. Guo, PhD, Human Genetics Center, Division of Biostatistics, The University of Texas School of Public Health; J. Wang, PhD, State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University; C. Shi, MD, Department of Rheumatology, Jiangxi People's Hospital; M. Espitia, BS, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; X. Guo, BS, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; Q. Wang, MD, PhD, Department of Rheumatology, Peking University-Shenzhen Hospital; M. Liu, PhD, Life Sciences College, Hubei University; S. Assassi, MD, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; J.D. Reveille, MD, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School; M.D. Mayes, MD, Division of Rheumatology, Department of Internal Medicine, University of Texas McGovern Medical School.

Address correspondence to Prof. X. Zhou, Department of Internal Medicine, Division of Rheumatology, The University of Texas Medical School at Houston, 6431 Fannin St., MSB5270, Houston, Texas 77030, USA. E-mail: Xiaodong.zhou@uth.tmc.edu

Accepted for publication August 8, 2018.

Systemic sclerosis (SSc) is an immune-mediated fibrotic disease with complex genetic features. It can be classified clinically by the extent of skin fibrosis into limited cutaneous (lcSSc) and diffuse cutaneous (dcSSc) forms[1]. Most patients with SSc (about 94%) have antinuclear autoantibodies (ANA)[2]. Some of these autoantibodies are SSc-specific, such as antibodies to topoisomerase I (ATA), centromere (ACA), and RNA polymerase III (ARA)[3]. Individual patients with SSc rarely have more than 1 of these autoantibodies[3]. Therefore, SSc can also be subgrouped by the presence of these autoantibodies, which are associated with specific clinical features[4].

Although the precise etiopathogenesis of SSc is still unknown, genetic predisposition is clearly an important factor. Multiple genetic loci have been associated with SSc[5]. Genome-wide association studies (GWAS) have indicated that the strongest SSc-associated loci fall within the HLA class II region[6,7]. Some specific alleles of classic HLA-II genes, such as *DRB1*01:01, DRB1*04:04, DRB*0701, DRB*11:04, DQA1*01:01, DQA*0201, DQA*0501, DQB*02:02, DQB1*05:01,* and *DPB*13:01*, have been established as SSc risk or protective factors[8]. In addition to these classic HLA-II genes, the SSc GWAS also showed that 2 synonymous single-nucleotide polymorphisms (SNP) in *NOTCH4*, a non-classical HLA gene in the HLA class II region, are associated with SSc[9].

NOTCH4 stands for neurogenic locus notch homolog protein 4, which is a member of the NOTCH family[10]. The NOTCH4 protein is a single-pass transmembrane receptor containing extracellular (NECD), transmembrane, and intracellular domains (NICD). The NECD contains 29-epidermal growth factor (EGF)-like repeats and serves for the ligands and calcium binding. Upon binding with ligands (including Jag 1 and 2, Delta-like 1, 3, and 4[11]), proteolysis of NOTCH occurs that releases NICD from the cell membrane. NICD is then translocated into the nucleus. NICD in turn activates expression of a group of downstream genes such as *Hes* and *Hay*[12]. NOTCH signaling is an essential pathway involving cell proliferation, differentiation, and apoptosis[13].

As a followup fine-mapping study of the GWAS, we sequenced whole *NOTCH4* exons, to identify NOTCH4 sequence changes that could contribute to the pathogenesis of SSc.

MATERIALS AND METHODS

Study subjects. A total of 1006 patients with SSc and 1004 controls of European ancestry were examined in the first cohort. A Chinese Han cohort composed of 576 patients and 574 controls was also examined to compare ethnic difference. All patients met either the 1980 American College of Rheumatology (ACR) criteria[14] or the 2013 ACR/European League Against Rheumatism criteria[15] for SSc. The ethical approval of the studies was obtained from the institutional review boards at The University of Texas McGovern Medical School (approval number: HSC-MS-10-0451). Informed written consent was obtained from each participant.

Autoantibodies tests. Patient's sera were tested for ANA by indirect immunofluorescence using HEp-2 cells as antigen substrate (Antibodies Inc.). ATA was detected by passive immunodiffusion against calf thymus extracts (Inova Diagnostics Inc.). ACA was determined by the pattern of indirect immunofluorescence using HEp-2 cells. ARA was detected using commercially available kits[16].

HLA-class II genotyping. HLA-DRB1, HLA-DQA1, HLA-DQB1, and HLA-DPB1 genotyping were performed with oligotyping[8] or sequence-based typing (SBT) method using SeCore Kits (Life Technologies)[17]. The HLA SBT uTYPE 6.0 program (Life Technologies) was used in the sequencing analysis and in assigning HLA alleles.

Next-generation sequence (NGS). Genomic DNA was extracted from the peripheral blood of each subject. DNA concentrations were measured by ABI Taman DNase P kit and 7900 real-time PCR machine (Applied Biosystems). The *NOTCH4* whole exonic sequencing was performed with the Ion Torrent Personal Genome Machine sequencer system (Applied Biosystems). The primer sets were designed by Ampliqseq designer 2.0 to cover the full length of exons and their franking intron regions at 100× coverage and 100% mapping rate. These were used for the sequence: Ion AmpliSeq Library Kit 2.0, Ion Library Equalizer kit, Ion 318 V2 Chip, Ion PGM template OT2 200 kit, and Ion PGM Sequencing 200 kit v2. The experiments were performed by following the protocols from the Ion Torrent manufacturer. Base calling, alignment, mapping, and variant calling were performed in Torrent Suite 4.0 with high stringency default setting. The human reference genome is HG19.

Confirmation of variants with Sanger sequencing (SS). SS was used to verify the findings from the NGS and to examine the Chinese cohorts. It was performed on an ABI3130xl Genetic Analyzer (Applied Biosystems), and data were analyzed using Sequencing Analysis v.5.2 software (Applied Biosystems). The success reading of SS was 100%.

Statistical analysis. The Hardy-Weinberg equilibrium (HWE) test was done in the control and SSc samples and the SNP were filtered out if the p value of HWE was < 0.05. Chi-square test was conducted to analyze the differences of variant counts between patients with SSc and controls. Exact p values (Fisher's test) were obtained from 2×2 tables of genotype counts and disease status if the sample size for any cells was < 5. Conditional logistic regression analysis was applied to eliminate risk- or protective-effects of known SSc-associated HLA alleles on *NOTCH4* polymorphisms. The p values of < 0.05 were considered statistically significant after the false discovery ratio (FDR)–based multiple test correction. Stratified association analyses were conducted with ATA, ACA, and lcSSc/dcSSc to check for an association of SNP in different subgroups. Haplotype analyses were conducted with Haploview software (4.2)[18]. All other association analyses were conducted in R with our custom scripts (3.2.0)[19] and Plink (1.07). CTG triplet repeat (located in chr6:32191659-32191690 in human hg19 genome) association analyses were conducted by comparing candidate alleles or genotypes with the most frequent allele or genotype in control samples.

RESULTS

NGS results and Sanger sequence verification in US cohort. Clinical and autoantibody information of patients with SSc are summarized in Supplementary Table 1 (available from the authors on request). A total of 143 exonic variants were reported from the NGS. Among them, 20 variants including 12 known and 8 new variants were associated with SSc by surpassing FDR-adjusted $p < 0.05$. Therefore, they were further examined with SS. However, all 8 new variants reported by the NGS appeared to be false variants as analyzed by SS. All 12 known SSc variants including 11 SNP and 1 triplet repeat (CTG, n = 5, 6, 9, 10, 11, 12, 13) were consistent between NGS and SS results, and met the HWE test (Supplementary Table 2, available from the authors on request).

The allele-based associations with SSc are shown in Table 1. In addition, dominant model, genotype model, recessive model, and trend model also were tested in the studies, and these are shown in Supplementary Table 3 (available from the authors on request).

Analysis for association between *NOTCH4* polymorphisms and clinical subtypes showed that the strongest associations with lcSSc and dcSSc were rs204987 ($p = 1.5 \times 10^{-7}$, OR = 0.23) and 3 linkage disequilibrium (LD)-linked SNP (rs520803/rs520692/rs520688; p = 0.0013, OR 1.35; Supplementary Table 4, available from the authors on request).

Table 1. Association of *NOTCH4* variants with SSc by allelic model in the US cohort.

SNP	Change	MAF (case/control)	p	OR (95% CI)	P Corrected
rs1044507	C/A	0.014/0.031	3.81×10^{-6}	0.38 (0.24–0.62)	4.57×10^{-5}
rs204987	T/C	0.042/0.063	1.1×10^{-7}	0.37 (0.24–0.56)	1.32×10^{-6}
rs415929	C/T	0.016/0.042	1.57×10^{-6}	1.34 (1.17–1.54)	1.89×10^{-5}
rs422951	C/T	0.326/0.278	4.02×10^{-4}	1.21 (1.06–1.38)	4.82×10^{-3}
rs423023	C/G	0.460/0.412	8.81×10^{-5}	1.27 (1.10–1.46)	1.06×10^{-3}
rs443198	G/A	0.344/0.284	1.26×10^{-5}	0.77 (0.67–0.88	1.51×10^{-4}
rs520688	C/T	0.344/0.284	5.44×10^{-6}	1.32 (1.15–1.51)	6.53×10^{-5}
rs520692	C/T	0.344/0.284	5.44×10^{-6}	1.32 (1.15–1.51)	6.53×10^{-5}
rs520803	T/C	0.344/0.285	5.44×10^{-6}	1.32 (1.15–1.51)	6.53×10^{-5}
rs8192579	C/T	0.319/0.362	3.75×10^{-4}	0.65 (0.49–0.88)	4.5×10^{-3}
rs915894	G/T	0.313/0.370	2.25×10^{-4}	0.81 (0.71–0.93)	2.7×10^{-3}
(CTG)n	10CTG	0.411/0.401	—	1.00	—
	12CTG	0.111/0.105	0.78	1.04 (0.84–1.28)	1.000
	11CTG	0.129/0.093	5.97×10^{-3}	1.35 (1.09–1.67)	0.036
	13CTG	0.011/0.029	4.97×10^{-5}	0.35 (0.21–0.59)	2.98×10^{-4}
	9CTG	0.203/0.271	1.66×10^{-4}	0.73 (0.62–0.86)	9.95×10^{-4}
	6CTG	0.132/0.099	0.015	1.3 (1.06–1.6)	0.093
	5CTG	0.004/0.003	0.83	1.3 (0.45–3.77)	1.000

SSc: systemic sclerosis; SNP: single-nucleotide polymorphism; MAF: minor allele frequency.

In analysis of SSc autoantibody subtypes, ATA+ patients with SSc were positively associated with 11CTG, 6CTG, and 6 SNP, including rs415929, rs520803/rs520692/rs520688, rs423023, and rs422951. ACA+ patients were negatively associated with 9CTG, 13CTG, and 5 SNP including rs443198, rs204987, rs8192579, rs915894, and rs1044507. Among them, the p values of 11CTG, rs415929, rs520803/rs520692/rs520688, and rs423023 passed the genome-wide significance thresholds of 5×10^{-8} (Table 2A). ARA+ patients with SSc did not show significant association with any variants of *NOTCH4*.

Because the *NOTCH4* gene is adjacent to HLA class II genes, potential LD from the latter may mask the association of *NOTCH4* with SSc. We therefore performed conditional logistical regression analysis based on the known SSc-associated HLA II genes including *DRB1*11:01* and *DPB1*13:01* for ATA+ SSc, and *DRB1*01:01, DRB1*04:04, DQA1*01:01*, and *DQB1*05:01* for ACA+ SSc[8]. All SNP, except rs422951 to ATA and rs1044507 to ACA, remained significantly associated with their corresponding autoantibody subtypes (Table 2B).

Association of NOTCH4 *variants with SSc in the Chinese cohort*. To investigate whether the associations found in US SSc varied by different ethnicity, a Chinese SSc cohort was examined for these 12 *NOTCH4* variants. SNP rs1044507 achieved a nominal significance threshold of $p < 0.05$ for association with SSc ($p = 6.8 \times 10^{-4}$, OR 0.56; Table 3). Allelic association of CTG tandems with SSc were found at 6CTG ($p = 3 \times 10^{-4}$, OR 1.67) and 9CTG ($p = 0.001$, OR 1.42). The former was in concordance with US cohort ($p = 0.015$, OR 1.3), while the latter was in the opposite direction of the US cohort.

For autoantibody subtypes, ATA+ SSc was in agreement with US cohort results positively associated with rs415929, rs520803, rs520692, rs520688, and rs423023. ACA+ SSc was positively associated with 6CTG, and negatively with rs1044507, which achieved nominal significance of $p < 0.05$, also consistent with the US cohort (Table 3).

The case numbers of subsets of Chinese SSc autoantibodies in specific SSc-associated HLA genotypes are too small for association analysis of NOTCH4 alleles to conclude the result.

In the Chinese cohort, the number of study subjects is limited. We cannot perform the analysis of the independent association of the NOTCH4 alleles from the HLA alleles. However, we performed LD analysis using 157 CHB (Han Chinese in Beijing) and CHS (Southern Han Chinese) samples obtained from the 1000 Genome Project. Tag-SNP were selected (between the GPX5 to the ZBTB9) to cover the whole HLA region after removing the redundant variants with high LD within 5 adjacent SNP (step = 3 and variance inflation factor = 2). LD was estimated between pairwise Tag-SNP and the distribution of the r^2 between *NOTCH4* and adjacent regions was calculated (Supplementary Figure 1, available from the authors on request). The result indicated that LD between the *NOTCH4* gene variants is very strong, but it becomes significantly decayed as its genomic distance extends beyond the *NOTCH4* genes in the HLA region in the Chinese population. The r^2 between the *NOTCH4* and HLA-DP, -DR, and –DQ appeared to be relatively weak ($r^2 < 0.045$).

DISCUSSION

Some HLA class II gene alleles, such as DRB1*0701, DRB1*11:04, DQA1*0201, DQA1*0501, DQB1*02:02,

Table 2. Association of *NOTCH4* variants with autoantibody subtypes of SSc in the US cohort.

A. Association without conditioning on SSc-associated HLA genotypes.

	ACA+ vs Control		ATA+ vs Control		ARA+ vs Control	
SNP	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
rs443198	0.54 (0.42–0.69)	1.6×10^{-6}	0.57 (0.42–0.77)	2.4×10^{-4}	0.77 (0.57–1.02)	0.069
rs915894	0.64 (0.50–0.83)	5.6×10^{-4}	0.61 (0.45–0.82)	0.001	0.69 (0.51–0.92)	0.012
rs415929	1.24 (0.97–1.57)	0.084	2.39 (1.82–3.13)	1.8×10^{-10}	1.38 (1.04–1.83)	0.024
rs520688	1.23 (0.97–1.57)	0.092	2.26 (1.72–2.97)	3.6×10^{-9}	1.35 (1.02–1.79)	0.039
rs520692	1.23 (0.97–1.57)	0.092	2.26 (1.72–2.97)	3.6×10^{-9}	1.35 (1.02–1.79)	0.039
rs520803	1.23 (0.97–1.58)	0.092	2.26 (1.72–2.97)	3.6×10^{-9}	1.35 (1.02–1.80)	0.039
rs422951	0.95 (0.75–1.21)	0.698	1.53 (1.15–2.02)	0.0029	1.01 (0.77–1.34)	0.935
rs423023	1.12 (0.87–1.45)	0.374	2.23 (1.67–2.97)	4.3×10^{-8}	1.24 (0.92–1.67)	0.151
rs204987	0.11 (0.03–0.35)	8×10^{-6}	0.45 (0.20–1.04)	0.057	0.66 (0.32–1.37)	0.263
rs8192579	0.31 (0.18–0.60)	2.5×10^{-4}	1.52 (0.92–2.51)	0.102	0.65 (0.35–1.23)	0.187
rs1044507	0.26 (0.10–0.65)	0.002	0.31 (0.10–0.91)	0.024	0.47 (0.19–1.18)	0.1
10CTG	Reference					
13CTG	0.24 (0.09–0.6)	1.6×10^{-3}	0.33 (0.1–1.05)	0.076	0.48 (0.19–1.21)	0.161
12CTG	0.67 (0.48–0.93)	0.022	0.57 (0.34–0.94)	0.036	0.66 (0.42–1.04)	0.088
11CTG	0.92 (0.67–1.26)	0.658	2.84 (2.07–3.9)	7.3×10^{-11}	1.4 (0.97–2.02)	0.085
9CTG	0.52 (0.41–0.67)	3.5×10^{-7}	0.95 (0.7–1.28)	0.783	0.87 (0.65–1.16)	0.372
6CTG	1.4 (1.06–1.84)	0.019	1.17 (0.79–1.74)	0.504	1.31 (0.91–1.89)	0.172
5CTG	0.92 (0.18–4.58)	1	1.05 (0.13–8.78)	1	1.85 (0.37–9.26)	0.788

B. Association under condition of SSc-associated HLA-class II alleles in the US cohort.

	CEN+ vs Control		ATA+ vs Control		ARA+ vs Control	
SNP	OR (L95–U95)	p	OR (L95–U95)	p	OR (L95–U95)	p
rs1044507	0.38 (0.12–1.17)	0.092	0.43 (0.2–0.82)	0.011	0.45 (0.22–0.88)	0.021
rs204987	0.13 (0.03–0.55)	0.0056	0.36 (0.20–0.65)	7×10^{-4}	0.33 (0.17–0.62)	6.5×10^{-4}
rs415929	1.15 (0.87–1.52)	0.329	1.34 (1.08–1.65)	0.0066	1.09 (0.88–1.36)	0.426
rs422951	0.91 (0.71–1.16)	0.431	1.15 (0.96–1.38)	0.14	1.06 (0.88–1.28)	0.549
rs423023	1.10 (0.83–1.47)	0.504	1.32 (1.07–1.63)	0.010	1.07 (0.86–1.34)	0.559
rs443198	0.56 (0.42–0.74)	4.1×10^{-5}	0.82 (0.67–0.99)	0.044	0.87 (0.71–1.07)	0.185
rs520688	1.15 (0.87–1.53)	0.322	1.34 (1.09–1.65)	0.006	1.10 (0.88–1.37)	0.389
rs520692	1.15 (0.87–1.53)	0.322	1.34 (1.09–1.65)	0.006	1.10 (0.88–1.37)	0.389
rs520803	1.15 (0.87–1.53)	0.322	1.34 (1.09–1.65)	0.006	1.10 (0.88–1.37)	0.389
rs8192579	0.30 (0.14–0.66)	0.0026	0.77 (0.50–1.19)	0.242	0.48 (0.29–0.78)	0.003
rs915894	0.72 (0.55–0.95)	0.018	0.83 (0.68–1.02)	0.077	0.89 (0.72–1.1)	0.276

SSc: systemic sclerosis; ACA: anticentromere antibodies; ATA: antibodies to topoisomerase I; ARA: anti-RNA polymerase III antibodies; SNP: single-nucleotide polymorphism; CEN: anticentromere autoantibodies.

DQB1*05:01, and DPB1*13:01, are established SSc risk or protective factors[8]. The genetic association of some HLA alleles appeared more significant in SSc autoantibody subsets. For instance, HLA-DRB1*11:04 and DPB1*13:01 confer strong risk of ATA-positive SSc, and DRB1*04:01 and DQB1*03:01 confer strong risk of SSc to ACA-positive patients[8]. The extensive LD across the HLA region has made it hard to distinguish true association between SSc and other HLA region genes or loci from the LD effect of these HLA alleles. We addressed this confounding issue in this post-GWAS sequencing study of the *NOTCH4* gene with conditional logistical regression analysis in a large SSc cohort of US white subjects, followed by analysis in a Chinese cohort that has different ancestral histories from US subjects and different LD patterns[17,20,21]. A total of 12 exonic variants of the *NOTCH4* gene were found to be associated with SSc in the US cohort. There was a stronger association of *NOTCH4* variants in the autoantibody subtypes of SSc. A CTG triplet repeat (11CTG) and 5 SNP (rs415929, rs520803/rs520692/rs520688, rs423023) located in the NECD of *NOTCH4* conferred a strong risk (p values passed genome-wide significance threshold) in ATA+ patients with SSc in the US cohort. The CTG triplets encode poly-leucines (number from 5 to 13), and the number of poly-leucines determines the length of the signal peptide domain, which then affects the binding of signal peptidase, and maturation and translocation of *NOTCH4*[10,11]. SNP rs520692 encodes a nonsynonymous polymorphism switching between aspartic acid and glycine. On the other hand, all identified ACA+ associations with *NOTCH4* variants (9CTG, 13CTG, and 5 SNP including

Table 3. Association of *NOTCH4* variants with SSc by allelic model in the Chinese cohort.

	SSc vs Control		ACA+ vs Control		ATA+ vs Control	
SNP	OR	p	OR	p	OR	p
rs1044507	0.56 (0.40–0.79)	6.8×10^{-4}	0.52 (0.29–0.95)	0.031	0.75 (0.29–1.89)	0.536
rs204987	2.50 (0.48–12.90)	0.258	2.04 (0.18–22.59)	0.552	0.00 (0.00–NaN)	0.713
rs415929	1.02 (0.84–1.25)	0.824	1.28 (0.94–1.74)	0.120	1.92 (1.16–3.17)	0.0099
rs422951	0.93 (0.76–1.13)	0.471	1.16 (0.85–1.58)	0.339	1.33 (0.79–2.24)	0.288
rs423023	1.03 (0.84–1.27)	0.774	1.27 (0.92–1.75)	0.139	1.75 (1.04–2.94)	0.033
rs443198	1.03 (0.88–1.22)	0.697	1.14 (0.87–1.48)	0.346	1.26 (0.79–2.02)	0.336
rs520688	1.03 (0.84–1.27)	0.788	1.28 (0.93–1.76)	0.127	1.78 (1.06–3.00)	0.028
rs520692	1.17 (0.93–1.47)	0.191	1.51 (1.08–2.12)	0.016	1.76 (1.00–3.11)	0.047
rs520803	1.16 (0.92–1.45)	0.213	1.49 (1.06–2.09)	0.020	1.74 (0.99–3.06)	0.053
rs8192579	1.07 (0.74–1.56)	0.719	1.20 (0.68–2.14)	0.525	1.67 (0.70–4.03)	0.245
rs915894	0.93 (0.78–1.09)	0.361	0.83 (0.64–1.09)	0.176	0.58 (0.35–0.95)	0.028
10CTG	Reference					
13CTG	0.7 (0.48–1.03)	0.086	0.73 (0.38–1.43)	0.449	0.71 (0.24–2.13)	0.718
12CTG	0.82 (0.23–2.93)	1	1.79 (0.35–9.1)	0.822	0 (0–∞)	1
11CTG	1.21 (0.92–1.6)	0.194	1.43 (0.92–2.23)	0.141	0.81 (0.35–1.88)	0.777
9CTG	1.42 (1.15–1.75)	0.001	1.37 (0.96–1.95)	0.099	1.19 (0.67–2.14)	0.651
6CTG	1.67 (1.27–2.19)	3×10^{-4}	2.05 (1.34–3.11)	0.001	1.21 (0.56–2.59)	0.773

SSc: systemic sclerosis; ACA: anticentromere antibodies; ATA: antibodies to topoisomerase I; SNP: single-nucleotide polymorphism.

rs443198, rs204987, rs8192579, rs915894, and rs1044507) were negative, except 6CTG. Among them, rs915894 encodes a nonsynonymous polymorphism switching between lysine and glutamine located in NECD. Importantly, most of these associations seemed independent from an LD effect of the HLA-class II genes that have been shown to be strongly associated with SSc[8]. It is worth noting that the synonymous SNP rs443198, previously reported in the SSc-GWAS[9], showed a consistent association with SSc here, especially with ACA+ patients.

Although not all associations of the *NOTCH4* variants with US SSc were concordant in the Chinese cohort, the strong negative association of rs1044507 and the positive association of 6CTG were consistent between the 2 cohorts. In addition, there was a noteworthy trend of association of the *NOTCH4* variants with SSc autoantibody subtypes. These findings support *NOTCH4* as an important genetic factor in SSc, and this is in agreement with the previous GWAS and family studies of SSc[9,22].

It is also worth noting that *NOTCH4* variants have been associated with other rheumatic and/or immune-mediated diseases including rheumatoid arthritis[23], sporadic inclusion body myositis[24], ulcerative colitis[25], alopecia areata[26], and age-related macular degeneration[27]. The CTG of *NOTCH4* form an α/β horseshoe fold on NECD believed to be involved in protein-protein interactions, which has also been implicated in patients with schizophrenia[28].

The functional significance of these *NOTCH4* variants remains to be elucidated. Several identified SSc-associated *NOTCH4* variants, especially nonsynonymous polymorphisms, are located in the NECD involved in construction of EGF-like repeats that may influence molecular structure and the binding of *NOTCH4* ligands and Ca2+. On the other hand, an activated *NOTCH* signaling has been reported in SSc[29] and other fibrotic disorders such as keloids and kidney fibrosis[30,31], and *NOTCH* signaling appeared to contribute to fibroblast activation and collagen overproduction in SSc and fibrotic animal models[29,32].

Several *NOTCH4* exonic variants were associated with SSc and/or SSc subtypes in studies of both US and Chinese cohorts. The results support the previous findings of a genetic contribution of *NOTCH4* to SSc. Our results indicated that these associations are independent of HLA II associations. This may provide novel insight into SSc-associated and specific changes of coding region sequence of the *NOTCH4* gene, some of which are nonsynonymous variants located in the NECD. The functional effect of these SSc-associated variants, especially on ligand binding and activation of *NOTCH4* signaling, need to be further explored.

REFERENCES

1. Medsger TA. Systemic sclerosis (scleroderma): clinical aspects. In: Koopman WJ, editor. Arthritis and allied conditions, 14th ed. Philadelphia: Lippincott Williams & Wilkins; 2000:1590.
2. Salazar GA, Assassi S, Wigley F, Hummers L, Varga J, Hinchcliff M, et al. Antinuclear antibody-negative systemic sclerosis. Semin Arthritis Rheum 2015;44:680-6.
3. Bunn CC, Black CM. Systemic sclerosis: an autoantibody mosaic. Clin Exp Immunol 1999;117:207.
4. 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.
5. Ma Y, Zhou X. Multiple genes, especially immune-regulating genes, contribute to disease susceptibility in systemic sclerosis. Curr Opin Rheumatol 2016;28:595-605.

6. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum 2009;60:3807-14.
7. 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.
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. Gorlova O, Martin J, Rueda B, Koeleman B, Ying J, Teruel M, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genetics 2011;7:e1002178.
10. Li L, Huang GM, Banta AB, Deng Y, Smith T, Dong P, et al. Cloning, characterization, and the complete 56.8-kilobase DNA sequence of the human NOTCH4 gene. Genomics 1998;51:45-58.
11. Baron M. An overview of the Notch signaling pathway. Semin Cell Dev Biol 2003;2:113-9.
12. Yabe Y, Matsumoto T, Tsurumoto T, Shindo H. Immunohistological localization of Notch receptors and their ligands Delta and Jagged in synovial tissues of rheumatoid arthritis. J Orthop Sci 2005;6:589-94.
13. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science 1999;5415:770-6.
14. Preliminary criteria for the classification of systemic sclerosis (scleroderma) Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23:581-90.
15. 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. Ann Rheum Dis 2013; 72:1747-55.
16. Zhou X, Tan FK, Xiong M, Milewicz DM, Feghali CA, Fritzler MJ, et al. Systemic sclerosis (scleroderma): specific autoantigen genes are selectively overexpressed in scleroderma fibroblasts. J Immunol 2001;167:7126-33.
17. He D, Wang J, Yi L, Guo X, Guo S, Guo G, et al. Association of the HLA-DRB1 with scleroderma in Chinese population. PLoS One 2014;9:e106939.
18. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.
19. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet. Accessed October 15, 2018.] Available from: www.R-project.org/
20. Wang J, Guo X, Yi L, Guo G, Tu W, Wu W, et al. Association of HLA-DPB1 with scleroderma and its clinical features in Chinese population. PLoS One 2014;9:e87363.
21. Zhou XD, Yi L, Guo XJ, Chen E, Zou HJ, Jin L, et al. Association of HLA-DQB1*0501 with scleroderma and its clinical features in Chinese population. Int J Immunopathol Pharmacol 2013; 26:747-51.
22. Cardinale CJ, Li D, Tian L, Connolly JJ, March ME, Hou C, et al. Association of a rare NOTCH4 coding variant with systemic sclerosis: a family-based whole exome sequencing study. BMC Musculoskelet Disord 2016;17:462.
23. Kochi Y, Yamada R, Kobayashi K, Takahashi A, Suzuki A, Sekine A, et al. Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. Arthritis Rheum 2004;50:63-71.
24. Scott AP, Laing NG, Mastaglia F, Dalakas M, Needham M, Allcock RJ. Investigation of NOTCH4 coding region polymorphisms in sporadic inclusion body myositis. J Neuroimmunol 2012;250:66-70.
25. Juyal G, Negi S, Sood A, Gupta A, Prasad P, Senapati S, et al. Genome-wide association scan in north Indians reveals three novel HLA-independent risk loci for ulcerative colitis. Gut 2015; 64:571-9.
26. Tazi-Ahnini R, Cork MJ, Wengraf D, Wilson AG, Gawkrodger DJ, Birch MP, et al. Notch4, a non-HLA gene in the MHC is strongly associated with the most severe form of alopecia areata. Hum Genet 2003;112:400-3.
27. Cipriani V, Leung HT, Plagnol V, Bunce C, Khan JC, Shahid H, et al. Genome-wide association study of age-related macular degeneration identifies associated variants in the TNXB-FKBPL-NOTCH4 region of chromosome 6p21.3. Hum Mol Genet 2012;21:4138-50.
28. Wassink TH, Nopoulos P, Pietila J, Crowe RR, Andreasen NC. NOTCH4 and the frontal lobe in schizophrenia. Am J Med Genet B Neuropsychiatr Genet 2003;118B:1-7.
29. Dees C, Tomcik M, Zerr P, Akhmetshina A, Horn A, Palumbo K, et al. Notch signalling regulates fibroblast activation and collagen release in systemic sclerosis. Ann Rheum Dis 2011;70:1304-10.
30. Kim JE, Lee JH, Jeong KH, Kim GM, Kang H. Notch intracellular domain expression in various skin fibroproliferative diseases. Ann Dermatol 2014;26:332-7.
31. Grabias BM, Konstantopoulos K. Notch4-dependent antagonism of canonical TGF-β1 signaling defines unique temporal fluctuations of SMAD3 activity in sheared proximal tubular epithelial cells. Am J Physiol Renal Physiol 2013;305:F123-33.
32. Dees C, Zerr P, Tomcik M, Beyer C, Horn A, Akhmetshina A, et al. Inhibition of Notch signaling prevents experimental fibrosis and induces regression of established fibrosis. Arthritis Rheum 2011;63:1396-404.