Accepted Manuscript

RXRB is a MHC-encoded susceptibility gene associated with anti-topoisomerase I antibody-positive systemic sclerosis

Akira Oka, Yoshihide Asano, Minoru Hasegawa, Manabu Fujimoto, Osamu Ishikawa, Masataka Kuwana, Yasushi Kawaguchi, Toshiyuki Yamamoto, Hiroki Takahashi, Daisuke Goto, Hirahito Endo, Masatoshi Jinnin, Shuhei Mano, Kazuyoshi Hosomichi, Tomotaka Mabuchi, Mahoko Takahashi Ueda, So Nakagawa, Stephan Beck, Seiamak Bahram, Kazuhiko Takehara, Shinichi Sato, Hironobu Ihn

PII: S0022-202X(17)31494-X

DOI: 10.1016/j.jid.2017.04.028

Reference: JID 860

To appear in: *The Journal of Investigative Dermatology*

Received Date: 25 January 2017

Revised Date: 14 April 2017

Accepted Date: 23 April 2017

Please cite this article as: Oka A, Asano Y, Hasegawa M, Fujimoto M, Ishikawa O, Kuwana M, Kawaguchi Y, Yamamoto T, Takahashi H, Goto D, Endo H, Jinnin M, Mano S, Hosomichi K, Mabuchi T, Ueda MT, Nakagawa S, Beck S, Bahram S, Takehara K, Sato S, Ihn H, *RXRB* is a MHC-encoded susceptibility gene associated with anti-topoisomerase I antibody-positive systemic sclerosis, *The Journal of Investigative Dermatology* (2017), doi: 10.1016/j.jid.2017.04.028.

RXRB is a MHC-encoded susceptibility gene associated with anti-topoisomerase I antibody-positive systemic sclerosis

Akira Oka[1]*, Yoshihide Asano[2], Minoru Hasegawa[3], Manabu Fujimoto[4], Osamu Ishikawa[5], Masataka Kuwana[6], Yasushi Kawaguchi[7], Toshiyuki Yamamoto [8], Hiroki Takahashi[9], Daisuke Goto[10], Hirahito Endo[11], Masatoshi Jinnin[12], Shuhei Mano[13], Kazuyoshi Hosomichi[14], Tomotaka Mabuchi[15], Mahoko Takahashi Ueda[16], So Nakagawa[16, 17], Stephan Beck[18], Seiamak Bahram[19], Kazuhiko Takehara[20], Shinichi Sato[2], Hironobu Ihn[12]

[1]The Institute of Medical Science, Tokai University, Kanagawa, Japan

[2]Department of Dermatology, University of Tokyo Graduate School of Medicine, Tokyo, Japan

[3]Department of Dermatology, School of Medicine, Faculty of Medical Sciences, University of Fukui, Fukui, Japan

[4]Department of Dermatology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

[5]Department of Dermatology, Gunma University Graduate School of Medicine, Gunma, Japan

[6]Department of Allergy and Rheumatology, Nippon Medical School Graduate School of Medicine, Tokyo, Japan

[7]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan

[8]Department of Dermatology, Fukushima Medical University, Fukushima, Japan

[9]Department of Rheumatology, Sapporo Medical University School of Medicine, Hokkaido, Japan

[10]Department of Internal Medicine, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

[11]Department of Rheumatology, Jusendo General Hospital, Fukushima, Japan

[12]Department of Dermatology and Plastic Surgery, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan

[13]Department of Mathematical Analysis and Statistical Inference, The Institute of Statistical Mathematics, Tokyo, Japan

[14]Department of Bioinformatics and Genomics, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan

[15]Department of Dermatology, Tokai University School of Medicine, Kanagawa, Japan

[16]Micro/Nano Technology Center, Tokai University, Kanagawa, Japan

[17]Department of Molecular Life Sciences, Tokai University School of Medicine, Kanagawa, Japan

[18]Medical Genomics, UCL Cancer Institute, University College London, London, UK

[19]INSERM UMR_S 1109, LabEx Transplantex, LIA France-Japan FJ-HLA, Centre de Recherche d'Immunologie et d'Hématologie, Faculté de Médecine, Strasbourg, France

[20]Department of Molecular Pathology of Skin, Faculty of Medicine, Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University, Kanazawa, Japan

*To whom correspondence should be addressed: 143 Shimokasuya, Isehara, Kanagawa, 259-1193, Japan. Tel: +81 463931121; Fax: +81 463964137

Email: oka246@is.icc.u-tokai.ac.jp

Short title: RXRB systemic sclerosis susceptibility gene

Abbreviations: RXRB, retinoid X receptor beta; ATA, anti-topoisomerase I antibody; MHC, major histocompatibility complex; SSc, systemic sclerosis; HLA, human leukocyte antigen; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; LD, linkage disequilibrium

ABSTRACT

Systemic sclerosis (SSc) is a systemic autoimmune and connective tissue disorder associated with the human leukocyte antigen (HLA) locus. However, the functional relationship between HLA gene(s) and disease development remains unknown. To elucidate major histocompatibility complex (MHC)-linked SSc genetics, we performed genotyping of MHC-borne microsatellites and *HLA-DPB1* alleles using DNA obtained from 318 anti-topoisomerase I antibody-positive patients and 561 healthy controls, all of Japanese descent. Those results revealed 2 MHC haplotypes associated with SSc. Exome sequencing and targeted analysis of these risk haplotypes identified rs17847931 in retinoid X receptor beta (*RXRB*) as a susceptibility variant (P=1.3×10^{-15}; odds ratio (OR)=9.4) with amino acid substitution p.V95A on the risk haplotype harboring *HLA-DPB1*13:01*. No identical variant in the other haplotype including *DPB1*09:01* was observed, though that haplotype also showed a significant association (P=8.5×10^{-22}; OR =4.3) with SSc. Furthermore, the number of risk factors was shown to be a predominant factor, as individuals with 2 factors had elevated risk (P=6.7 × 10^{-13}; OR=30.2). We concluded that *RXRB* may be involved in anti-fibrotic activity in skin and chromatin remodeling.

INTRODUCTION

Systemic sclerosis (SSc, MIM #181750) is a systemic autoimmune and connective tissue disorder (Nikpour et al., 2010) resulting in clinically heterogeneous symptoms that range from limited skin involvement to diffuse skin sclerosis with severe internal organ involvement (Jimenez and Derk, 2004). SSc patients are classified into 2 clinical subgroups; limited cutaneous SSc (lcSSc) and diffuse cutaneous SSc (dcSSc), with those in the latter associated with distinct clinical complications and prognosis (Gorlova et al., 2011). SSc patients carry different auto-antibodies, including anti-DNA topoisomerase I (ATA) and anti-centromere auto-antibodies (LeRoy and Medsger, 2001), with ATA found in approximately 40% of all patients with SSc (Hasegawa et al., 2013), though more frequently in those with dcSSc as compared to lcSSc (Hasegawa et al., 2013). Furthermore, SSc occurs significantly more frequently in families with scleroderma than in the general population (Arnett et al., 2001).

Several autoimmune diseases, such as rheumatoid arthritis are genetically associated with alleles of human leukocyte antigen (HLA) genes within the major histocompatibility complex (MHC) on chromosome 6p21.3. Also, associations between SSc and HLA genes, especially class II, have been observed in various ethnic groups (Kuwana et al., 1999; Arnett et al., 2010). However, those studies did not elucidate whether HLA genes themselves or other MHC-embedded genes are SSc susceptibility factors. Genome-wide association studies of SSc performed in Europe also indicated that the strongest association with SSc is located in the HLA class II region (Allanore et al., 2011; Gorlova et al., 2011), though they did not identify the actual MHC-linked SSc susceptibility gene. Therefore, the genetic architecture of SSc in the MHC region remains unclear.

The genetic nature of the MHC region shows that multiple haplotypes with the highest degree of diversity are often maintained in the population by balancing selection, with positive selection occasionally generating long-range haplotypes. Thus, the strong linkage

disequilibrium (LD) observed even between distant loci within MHC can mask discrimination between a *bona fide* variant associated with disease and a variant influenced by LD (Kawashima et al., 2012; de Bakker et al., 2006). Microsatellites have higher mutation rates than single nucleotide polymorphisms (SNPs), which can result in break-up of apparently immutable common SNP haplotypes into lower frequency haplotypes, thereby facilitating identification of functional variants with intermediate or even rare frequencies (Gulcher, 2012). Therefore, multi-allelic microsatellites may be effective for pinpointing rare disease-associated haplotypes in the MHC.

Using a strategy based on these factors, we designed the present study with specific steps to identify SSc susceptibility variants. First, we performed association analysis using microsatellites for the MHC region with DNA obtained from Japanese SSc patients with ATA and healthy controls, and identified haplotypes associated with SSc. Next, we sequenced all exons of the MHC region into SSc candidate haplotypes (selected subjects with risk haplotype) and several control haplotypes (selected subjects without risk), then extracted variants that were included in only identical risk haplotypes from all variants detected by this sequencing method. Finally, we genotyped the candidate variants with identical risk haplotypes in all subjects, and re-estimated the haplotypes between the variants and microsatellites. Based on this result, we evaluated which locus generated haplotype associations with SSc, in other words, which locus contributes to elevate the frequencies of risk haplotypes in SSc patients. Consequently, we identified a SSc susceptibility variant (rs17847931) in retinoid X receptor beta (*RXRB*), a gene thought to be involved in anti-fibrotic activity in the skin as well as chromatin remodeling.

RESULTS

Association Analysis Using Multi-Allelic Loci within MHC Region

In all microsatellites, allele 171 of *D6S1583* showed the strongest association with SSc [odds ratio (OR) = 4.06, 95% confidence interval (CI) = 3.03–5.44, $P = 8.66\times10^{-23}$] and allele 162 of *D6S1701* showed the highest OR with statistical significance (OR = 5.08, 95% CI = 2.71–9.51, $P = 9.60\times10^{-08}$) (Figure 1a, b; see Supplementary Table S1 online). Moreover, *D6S2731* and *D6S1701* contained 2 alleles that showed strong associations with SSc (Figure 1a, b; see Supplementary Table S1 online). These loci were located in the HLA class II region. Among the *HLA* genes, these loci were the closest to *HLA-DPB1*, which also showed a significant association with SSc (Zhou et al., 2009; Arnett et al., 2010; Kuwana et al., 1999). Thus, we genotyped *HLA-DPB1* in the same subjects in addition to these microsatellites, which demonstrated that the alleles *DPB1*09:01* (OR = 3.79, 95% CI = 2.85–5.02, $P = 1.86\times10^{-22}$) and **13:01* (OR = 7.40, 95% CI = 4.13–13.2, $P = 6.80\times10^{-15}$) each had a significant association with SSc (see Supplementary Table S2 online).

Associations between SSc and Haplotypes Defined by Multi-allelic Loci

Haplotype analysis is essential for genetic studies within the MHC region, as that exhibits a higher level of diversity and more extended LD than any other loci (de Bakker et al., 2006). In the present study, we detected multiple alleles that showed significant associations with SSc in a single locus. Therefore, we speculated that *DPB1*09:01* and **13:01* are harbored in different haplotypes showing an association with SSc. First, we attempted to extract 2 alleles showing LD to *DPB1*09:01* and **13:01* in each multi-allelic locus, for which the association between SSc and alleles that showed the highest D′ value for *HLA-DPB1*09:01* (association 1) or **13:01* (association 2) in each locus was evaluated (Figure 1a, b; see Supplementary Table S1 online). These results suggested that at least 2 haplotypes associated with SSc are

harbored in the HLA class II region.

To define haplotypes associated with SSc, we explored segments that exhibited strong LD (r^2 >0.7) to *DPB1*09:01* or **13:01*, which indicated that one (association 1) was present at 6 loci (*D6S1100i*, *DPB1*, *D6S0512i*, *D6S2731*, *D6S1701*, *D6S1583*), while the other (association 2) was present at 2 loci (*DPB1*, *D6S2731*) (Figure 1c, d). Moreover, we estimated the haplotypes for 6 loci using PHASE v2.1.1 (Stephens and Donnelly, 2003). Only 2 haplotypes showed significant associations with SSc among approximately 400 examined (including rare haplotypes), both of which harbored *DPB1*09:01* [risk haplotype 1 (RH1)] and **13:01* (RH2), as expected (Table 1).

Sequencing of Risk Haplotypes

Next, we determined whether SSc susceptibility variants were harbored in the SSc risk haplotypes. For this, we sequenced 13 non-risk, 9 RH1, and 12 RH2 chromosomes (haplotypes) in the patient samples using whole-exome sequencing (see Supplementary Table S3 online). Among 5077 variants detected within the MHC region, we attempted to detect variants that were identical to each risk haplotype, but not observed in non-risk haplotypes. As a result, 37 variants were detected in RH1 and 11 in RH2 (Figure 1a, b; see Supplementary Figure S1 online), all of which were single nucleotide variants (SNVs). Based on pairwise LD analysis of 89 Japanese individuals obtained from the 1000 Genomes Browser (http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/), we also confirmed that 37 SNVs and 11 SNVs exhibited strong LD in 1702 kb (chr 6: 31080320–32782897) (RH1) and 124 kb (chr 6: 33048348–33172755) (RH2) (Figure 1a, b; see Supplementary Figure 2, 3 online). Among these, 10 and 2 SNVs conferred amino acid substitutions in RH1 and RH2, respectively (Table 2, see Supplementary Figure S1 online). Six non-synonymous SNVs that mutually exhibited strong LD in *BTNL2* were considered to be candidates for SSc

susceptibility in RH1, however, it is unlikely that these are related to a predisposition towards SSc, because the allele frequencies in Japanese are actually similar with the frequency of RH1 in the patient subjects enrolled in this study (Table 1, 2). Moreover, the DPB1-F64Y substitution produced by *allele A* of *rs1042117* is unique among all *DPB1* alleles observed in a Japanese population, whereas in other ethnic groups this substitution is not specific for *DPB1*13:01*, but is instead shared with several *DPB1* alleles (see Supplementary Table S4 online). As a result, we estimated that 4 variants (*rs2270191, rs117951780, rs1063646, rs79517313*) were candidate loci for SSc in RH1 and 1 (*rs17847931*) in RH2.

Identification of Susceptibility Variant(s)

Next, to confirm that the SNVs were harbored within each risk haplotype, candidate SNVs of all patients and controls were genotyped using the Sanger sequencing method. We selected *rs117951780* located in *CDSN* to represent the 4 variants that exhibited strong LD in RH1 as well as rs17847931 located in *RXRB* in RH2 (Table 2, see Supplementary Figure 2, 3 online). Both of these non-synonymous SNVs were significantly associated with SSc (*rs117951780*: OR = 2.66, 95% CI = 2.04–3.47, *P* = 1.06E-13; rs17847931: OR = 9.44, 95% CI = 4.97–17.9, *P* = 1.30E-15) (see Supplementary Table S5 online). We evaluated these results based on a haplotype analysis technique that included the SNVs and microsatellites to determine whether both SNVs predisposed carriers toward SSc, which indicated that *allele A* of *rs117951780* was not a primary factor for SSc in RH1, because a haplotype with allele G was also found to be associated with SSc (Table 3). Thus, the other 3 SNVs (*rs2270191, rs1063646, rs79517313*) that exhibited strong LD with rs117951780 (see Supplementary Figure 2, 3 online) were also not considered to be related to SSc susceptibility. As a result, a susceptibility variant in RH1 was not identified in this study. In contrast, of all haplotypes in RH2, one haplotype with *allele C* of *rs17847931* was found to be significantly associated

with SSc (OR = 9.44, 95% CI = 4.97–17.9, *P* = 1.30E-15) (Table 3), and was included with *DPB1*13:01* and *allele 214* of *D6S2731* (Table 3). While *allele C* of *rs17847931* was shown to be a candidate factor of SSc susceptibility, *DPB1*13:01* and *allele 214* of *D6S2731* were also significantly associated with SSc. Accordingly, this analysis could not identify which allele on this haplotype predisposes individuals to SSc. However, the haplotype with *allele C* of *rs17847931* was shown to never be combined with any alleles except for *DPB1*13:01* and *allele 214* of *D6S2731* (Figure 2, Table 3). Conversely, haplotypes harboring *DPB1*13:01* and/or *allele 214* of *D6S2731* combined with both alleles for *rs17847931* (Figure 2, Table 3). Moreover, no recombinant was observed in any haplotype that harbored *allele C* of *rs17847931* within the 3 loci *DPB1*, *rs17847931*, and *D6S2731*, though several recombinants in haplotypes with *allele T* of *rs17847931* were detected (Figure 2). Thus, *allele C* of *rs17847931* recently emerged in the haplotype that includes *DPB1*13:01* and *allele 214* of *D6S2731*, and may have an increased frequency with the haplotype. Furthermore, alleles of loci adjacent to rs17847931 also exhibited increased frequency in the patient group, possibly due to hitchhiking (Figure 1, see Supplementary Table S1 online). We cannot completely exclude the possibility of intergenic variants with SSc susceptibility based on the results of this study. However, the OR value for rs17847931 identified by exome sequencing was approximately equal to the RH2 value defined mainly by intergenic markers (Table 1 and 3, Figure 2). If an intergenic variant associated with SSc was within the RH2 value, then the risk would be lower as compared to that for RH2. Thus, there were no intragenic variants with stronger risk than the susceptibility variant rs17847931 within RH2. As a result, we concluded that *allele C* of *rs17847931* in *RXRB* is the primary allele associated with SSc.

Association Between Number of Risk Factors and SSc

To better understand how RH1 and *allele C* of *rs17847931* may influence an individual's risk

for SSc, we investigated the associations based on diplotypes and number of risk factors. Twenty of the SSc patients had both risk factors, whereas individuals who harbored both risk factors were never detected in the control group (see Supplementary Table S6 online), implying that these are strongly related to a predisposition for SSc. Our analysis based on the number of risk factors indicated that >50% of SSc patients had 1 or more risk factors (OR = 6.16, 95% CI = 4.44–8.60, *P* = 3.83E-32), while patients with 2 risk factors had higher OR values (OR = 30.2, 95% CI = 7.17–127, *P* = 6.73E-13) (Table 4). These results demonstrated that the number of risk factors within the MHC region is strongly related to SSc, indicating that other factors in RH1 leading towards a predisposition for SSc may be discovered in future studies.

Evaluation of impact *allele C* of *rs17847931* on RXRB function using *in silico* analysis

RXRs are mainly comprised of three domains; the N-terminal domain (NTD), where an amino acid with substitution by the *allele C* of *rs17847931* is located, the DNA binding domain (DBD), and the C-terminal ligand binding domain (LBD). The NTD is highly variable in terms of size and sequence among RXRs, and lacks a fixed tertiary structure, thus is termed an intrinsically disordered region. The NTD provides interaction surfaces for transcriptional co-regulatory proteins and protein binding to NTD mediates transcriptional activation of RXR receptors after undergoing a disorder-to-order transition (Kumar and McEwan, 2012).

To understand the influence of the substitution p.V95A (c.284T>C) in *RXRB*, we first defined the intrinsically disordered regions in the amino acid sequence of RXRB using several methods available in PONDR (Romero et al., 2001) and DISOPRED3 (Ward et al., 2004). All methods predicted that residue 95 was in the disordered region (see Supplementary Figure S4 online). Next, disordered protein-binding regions of the sequence were predicted

using Anchor (Meszaros, Simon and Dosztanyi, 2009; Dosztanyi, Meszaros and Simon, 2009), which showed that the residue at position 95 was located at the edge of the protein binding region from 67 to 95 (see Supplementary Table S7 online).

DISCUSSION

Previous studies have indicated that the *DPB1* locus is a susceptibility gene for ATA-positive SSc within the MHC region, and that *DPB1*13:01* is strongly associated with SSc and ATA in various ethnic groups (Arnett et al., 2010; Wang et al., 2014; Zhou et al., 2009). In our previous study and that of another group, a significant association of SSc with ATA and *DPB1*13:01*, as well as *DPB1*09:01* was shown in Japanese and Korean subjects (Kuwana et al., 1999; Zhou et al., 2009). Other studies have also reported these associations, though they did not determine whether *HLA* or other genes generated them. In the present study, we used microsatellite mapping combined with risk haplotype sequencing and clearly defined 2 genetic risk haplotypes, and also discovered at least 1 MHC-linked susceptibility variant (rs17847931) in *RXRB* related to ATA-positive SSc.

The present results identified a factor that predisposes individuals toward SSc in RH1, which covers approximately 1.7 Mbp (from *HLA-C* to *DPB1*), at least within the MHC region. Due to exome sequencing, we might have missed a susceptibility variant for SSc in RH1 if that was located in intergenic or intronic segments. However, a haplotype itself is more likely to predispose toward SSc. The haplotype harboring *DPB1*09:01* is a common long-range haplotype that spans the entire HLA class I and II regions, and is specific to the Japanese population (Okada et al., 2015). In addition, it is unlikely that a variant alone can protect a long-range haplotype including itself from recombination events without breaking down and increasing the haplotype frequency. Therefore, combinations of amino acid polymorphisms in multiple class I and II HLA genes may explain the risk factor for RH1 within the MHC region

(Okada et al., 2015).

RXRB is 1 of 3 retinoid X receptors (RXRs) belonging to the class of nuclear receptors, and mediates the effects of retinoic acid (RA) (Fitzgibbon et al., 1993). This receptor is a transcription factor that mediates an array of extracellular signals in a ligand-dependent manner to regulate the target gene by binding to response elements within the promoter regions of these genes (Philip et al., 2012). RXRs form homodimers or heterodimers with another nuclear receptor to bind RA and vitamin D (Evans and Mangelsdorf, 2014). Among retinoids, 9-cis RA is a high affinity ligand for RXRs (Heyman et al., 1992) and also exhibits a high anti-fibrotic activity in SSc fibroblasts (Xiao et al., 2011; Xiao et al., 2008). Homodimer formation by RXRs is induced by 9-cis RA (Zhang et al., 1992). RXRs also form a heterodimer with peroxisome proliferator-activated receptors (PPARs) and 9-cis RA regulates target genes via the PPAR/RXR complex (Kliewer et al., 1992). PPAR-γ also exhibits anti-fibrotic activity (Wu et al., 2009), and its mRNA/protein expression levels are reduced in SSc skin biopsy as well as in explanted skin fibroblast specimens (Dantas et al., 2015; Wei et al., 2010). Thus, RXRB likely exhibits anti-fibrotic activity in skin tissues via agonists such as RA.

ATP-driven chromatin remodeling activities and co-activators that exhibit intrinsic histone acetyltransferase activities are both involved in transcriptional initiation triggered by liganded heterodimeric RAR/RXR (Dilworth et al., 2000; Dilworth et al., 1999), while thyroid hormone receptor-β/RXRα heterodimers also induce localized disruption of chromatin in a thyroid hormone-independent manner (Lee et al., 2003). Thus, RXRB is also expected to be shown to be involved with histone acetylation and chromatin remodeling. The histone deacetylase inhibitor trichostatin A has an anti-fibrogenic effect in models of bleomycin-induced fibrosis (Ye et al., 2014; Huber et al., 2007) and also in SSc fibroblasts

(Wang, Fan and Kahaleh, 2006), implying that histone acetylation induces such anti-fibrotic action. Histone acetylation involved with RXR may also induce anti-fibrotic action.

Results of our *in silico* analysis suggested that the p.V95A substitution may alter the interaction of RXRB with the co-regulatory protein. Since hydrophobicity is important for protein-protein interaction, its decrease from V (very strong) to A (weak) may have a notable affect on protein-protein binding affinity, which has an impact on RXRB functions. Therefore, the function of this variant in *RXRB* should be elucidated by functional analysis in the future.

To summarize, we genetically identified *RXRB* as a gene with a high risk for SSc susceptibility, which is involved with anti-fibrotic activity in skin tissue. Our findings may lead to new therapeutic targets; indeed, drugs targeting RXRs are among the most widely used and commercially successful (Evans and Mangelsdorf, 2014). Future studies of RXRB may identify unknown aspects of the pathogenesis of SSc, thereby facilitating development of new therapeutic strategies and targets.

MATERIALS AND METHODS

Study approval

Upon the approval of the experimental procedures by relevant ethical committees at all universities, participants gave written informed consent in accordance with the Declaration of Helsinki.

Study population

We used unrelated ATA positive SSc patients (the age of onset, median [25 - 75 percentiles]: 48 years [34 - 59]), including 260 with diffuse cutaneous involvement and 252 with interstitial lung disease, and healthy unrelated individuals. A total of 318 individuals affected with SSc and 561 unrelated healthy individuals of Japanese origin participated in this

study. All the patients fulfilled the new classification criteria (van den Hoogen et al., 2013).

Microsatellite Genotyping

We selected 15 microsatellites spanning 5.81 Mbp in the HLA region (see Supplementary Table S8 online), and performed genotyping for patients and control subjects. All of the physical positions of the microsatellites and other loci on chromosome 6 were based on the reference UCSC Genome Browser assembly (GRCh37/hg19, http://genome.ucsc.edu/). Forward primers in the primer sets used to amplify microsatellites were labeled by 5' fluorescent FAM. Fragment analysis by capillary electrophoresis using a Thermo Fisher Scientific 3730 Genetic Analyzer and allele assignment using GeneMapper Software (Thermo Fisher Scientific) were performed as previously described (Tamiya et al., 2005). Fragment size was assigned to an allele name in corresponding microsatellites.

HLA-DPB1 **and Candidate Variants Genotyping**

Exon 2 of *HLA-DPB1*, and 2 variants of the *CDSN* (rs117951780) and *RXRB* (rs17847931) genes were sequenced using PCR-based Sanger sequencing. PCR assays were performed using a reaction volume of 25 μL, which contained 25 ng of genomic DNA, 0.5 U of KOD FX Neo (TOYOBO), 12.5 μL of 2× PCR Buffer, 5 μL of dNTP (2 mM each), and 0.2 μM (final concentration) of each of the primers. PCR sequencing primers and the thermal cycling profile are indicated in Supplementary Table S8 online. The PCR products were purified using AMPure XP (Beckman Coulter), according to the manufacturer's protocol. Purification and sequencing of the PCR products were conducted using a BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific) with a sequencing primer (see Supplementary Table S9 online) and a BigDye XTerminator Purification Kit (Thermo Fisher Scientific), according to the manufacturer's instructions. Automated electrophoresis was performed using

an ABI PRISM 3730 Genetic Analyzer (Thermo Fisher Scientific). DPB1 alleles were determined based on the alignment database of dbMHC.

Genomic Library Construction and Sequencing

We selected 17 subjects for exome sequencing based on haplotypes within the MHC region (see Supplementary Table S3 online). For exon fragment capture and sequencing, we used Agilent SureSelect Target Enrichment (v.5; 50 Mbp), according to the manufacturer's instructions. Sequence analysis was performed using Illumina Genome Analyzer IIx or HiSeq2000 platforms with the paired-end sequencing protocol (2×100 bp).

Analysis of Exome Sequencing Data in MHC Region

Fastx-toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) was used for quality control of the sequencing reads. Quality passed reads were mapped to the reference (UCSC Genome Browser assembly GRCh37/hg19, http://genome.ucsc.edu/) using Burrows-Wheeler Aligner (BWA) version 0.5.9 with the default parameters (Li and Durbin, 2009). After alignment, Sequence Alignment/Map (SAMtools) version 0.1.17 was used to convert the .sam file to a .bam file (Li et al., 2009), and potential PCR duplicates were flagged with Picard MarkDuplicates (version 1.88; http://picard.sourceforge.net/). Genome Analysis Toolkit (GATK, version 2.2-8) was used to perform local realignment, map quality score recalibration, and variant detection (McKenna et al., 2010). SNVs and indels were then annotated for functional consequences at gene and protein sequence levels using ANNOVAR (Wang, Li and Hakonarson, 2010). In the MHC region, the target of this study (chr6:25000001-35000000, hg19), samples from 17 selected subjects (see Supplementary Table S3 online) were sequenced to a mean depth of 65.1 reads, covering a mean 99.9% and 93.5% by at least 1 and 10 reads, respectively. Functions of non-synonymous SNVs were

predicted using SIFT (Kumar, Henikoff and Ng, 2009), PolyPhen 2 (Adzhubei et al., 2010), and FATHMM (Shihab et al., 2014).

Statistical Analysis

Logistic regression models were used to assess the genetic effects of multi-allelic loci and SSc risk haplotypes. Comparisons of differences in genotype and haplotype frequencies were performed using regression analysis for log-additive models (de Cid et al., 2009). The unadjusted OR and 95% CI values were calculated. Analyses were conducted using the SNPassoc R library (de Cid et al., 2009) obtained from The Comprehensive R Archive Network. For these association analyses, we used Bonferroni-corrected values to address the problem of multiple testing with a threshold *P*-value of 3.07E-04 (to account for multiple testing of 163 alleles in 16 multi-allelic loci in the first microsatellite analysis). The exact *P*-value for the Hardy–Weinberg proportion test was simulated using the Markov chain method within Genepop (Rousset, 2008). PHASE v2.1.1 was used to estimate haplotypes for multi-allelic loci (Stephens and Donnelly, 2003). To evaluate the LD around the MHC region, each multi-allelic locus was regarded as an SNV (Kawashima et al., 2012). Thus, we extracted the allele with the highest D′ value for 2 SSc risk alleles (*HLA-DPB1*09:01* and **13:01*) from all of the alleles in each locus, and merged the other alleles. Haploview 4.2 was used to investigate the degree of LD for risk haplotypes (Barrett et al., 2005). *P*-values (Fisher's exact test) and ORs were calculated for the association with SSc according to diplotype and haplotype numbers using the basic tool in the Comprehensive R Archive Network.

Analysis of RXRB N-terminal domain

To conduct *in silico* analysis of disordered regions in RXRB, we obtained the amino acid sequence (GenBank accession: NM_021976) from the NCBI GenBank database. To predict intrinsic disorder regions in the sequence, we used two computational programs; PONDR (http://www.pondr.com) and DISOPRED3 (Ward et al., 2004). For prediction by PONDR, we used four different methods that are implemented within the program (Peng et al., 2005; Obradovic et al., 2005; Obradovic et al., 2003; Romero et al., 2001; Romero, Obradovic and Dunker, 1997) To predict protein-binding regions, we utilized Anchor (Meszaros, Simon and Dosztanyi, 2009; Dosztanyi, Meszaros and Simon, 2009).

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare in regard to this study.

ACKNOWLEDGMENTS

This study was supported by a Research on Intractable Diseases grant from the Ministry of Health, Labor, and Welfare of Japan. We are grateful for Wang Ting, Hisako Kawada, Masayuki Tanaka, and Hideki Hayashi, as well as the Support Center for Medical Research and Education, Tokai University.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010 Apr;7(4):248-249.

Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet 2011 Jul;7(7):e1002091.

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 Jun;44(6):1359-1362.

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 May;69(5):822-827.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005 Jan 15;21(2):263-265.

Dantas AT, Pereira MC, de Melo Rego MJ, da Rocha LF,Jr, Pitta Ida R, Marques CD, et al. The Role of PPAR Gamma in Systemic Sclerosis. PPAR Res 2015;2015:124624.

de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006 Oct;38(10):1166-1172.

de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nat Genet 2009 Feb;41(2):211-215.

Dilworth FJ, Fromental-Ramain C, Remboutsika E, Benecke A, Chambon P. Ligand-dependent activation of transcription in vitro by retinoic acid receptor alpha/retinoid X receptor alpha heterodimers that mimics transactivation by retinoids in vivo. Proc Natl Acad Sci U S A 1999 Mar 2;96(5):1995-2000.

Dilworth FJ, Fromental-Ramain C, Yamamoto K, Chambon P. ATP-driven chromatin remodeling activity and histone acetyltransferases act sequentially during transactivation by RAR/RXR In vitro. Mol Cell 2000 Nov;6(5):1049-1058.

Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 2009 Oct 15;25(20):2745-2746.

Evans RM, Mangelsdorf DJ. Nuclear Receptors, RXR, and the Big Bang. Cell 2014 Mar 27;157(1):255-266.

Fitzgibbon J, Gillett GT, Woodward KJ, Boyle JM, Wolfe J, Povey S. Mapping of RXRB to human chromosome 6p21.3. Ann Hum Genet 1993 Jul;57(Pt 3):203-209.

Gorlova O, Martin JE, Rueda B, Koeleman BP, 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 Genet 2011 Jul;7(7):e1002178.

Gulcher J. Microsatellite markers for linkage and association studies. Cold Spring Harb Protoc 2012 Apr 1;2012(4):425-432.

Hasegawa M, Imura-Kumada S, Matsushita T, Hamaguchi Y, Fujimoto M, Takehara K. Anti-topoisomerase I antibody levels as serum markers of skin sclerosis in systemic sclerosis. J Dermatol 2013 Feb;40(2):89-93.

Heyman RA, Mangelsdorf DJ, Dyck JA, Stein RB, Eichele G, Evans RM, et al. 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. Cell 1992 Jan 24;68(2):397-406.

Huber LC, Distler JH, Moritz F, Hemmatazad H, Hauser T, Michel BA, et al. Trichostatin A prevents the accumulation of extracellular matrix in a mouse model of bleomycin-induced skin fibrosis. Arthritis Rheum 2007 Aug;56(8):2755-2764.

Jimenez SA, Derk CT. Following the molecular pathways toward an understanding of the pathogenesis of systemic sclerosis. Ann Intern Med 2004 Jan 6;140(1):37-50.

Kawashima M, Ohashi J, Nishida N, Tokunaga K. Evolutionary analysis of classical HLA class I and II genes suggests that recent positive selection acted on DPB1*04:01 in Japanese population. PLoS One 2012;7(10):e46806.

Kliewer SA, Umesono K, Noonan DJ, Heyman RA, Evans RM. Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. Nature 1992 Aug 27;358(6389):771-774.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-1081.

Kumar R, McEwan IJ. Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. Endocr Rev 2012 Apr;33(2):271-299.

Kuwana M, Inoko H, Kameda H, Nojima T, Sato S, Nakamura K, et al. Association of human leukocyte antigen class II genes with autoantibody profiles, but not with disease susceptibility in Japanese patients with systemic sclerosis. Intern Med 1999 Apr;38(4):336-344.

Lee KC, Li J, Cole PA, Wong J, Kraus WL. Transcriptional activation by thyroid hormone receptor-beta involves chromatin remodeling, histone acetylation, and synergistic stimulation by p300 and steroid receptor coactivators. Mol Endocrinol 2003 May;17(5):908-922.

LeRoy EC, Medsger TA,Jr. Criteria for the classification of early systemic sclerosis. J Rheumatol 2001 Jul;28(7):1573-1576.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009 Jul 15;25(14):1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009 Aug 15;25(16):2078-2079.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010 Sep;20(9):1297-1303.

Meszaros B, Simon I, Dosztanyi Z. Prediction of protein binding regions in disordered proteins. PLoS Comput Biol 2009 May;5(5):e1000376.

Nikpour M, Stevens WM, Herrick AL, Proudman SM. Epidemiology of systemic sclerosis. Best Pract Res Clin Rheumatol 2010 Dec;24(6):857-869.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53 Suppl 6:566-572.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2005;61 Suppl 7:176-182.

Okada Y, Momozawa Y, Ashikawa K, Kanai M, Matsuda K, Kamatani Y, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat Genet 2015 Jul;47(7):798-802.

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 2005 Feb;3(1):35-60.

Philip S, Castro LF, da Fonseca RR, Reis-Henriques MA, Vasconcelos V, Santos MM, et al. Adaptive evolution of the Retinoid X receptor in vertebrates. Genomics 2012 Feb;99(2):81-89.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001 Jan 1;42(1):38-48.

Romero, Obradovic, Dunker K. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome Inform Ser Workshop Genome Inform 1997;8:110-124.

Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 2008 Jan;8(1):103-106.

Shihab HA, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum Genomics 2014 Jun 30;8:11-7364-8-11.

Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003 Nov;73(5):1162-1169.

Tamiya G, Shinya M, Imanishi T, Ikuta T, Makino S, Okamoto K, et al. Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. Hum Mol Genet 2005 Aug 15;14(16):2305-2321.

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 Nov;72(11):1747-1755.

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 Jan 31;9(1):e87363.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010 Sep;38(16):e164.

Wang Y, Fan PS, Kahaleh B. Association between enhanced type I collagen expression and epigenetic repression of the FLI1 gene in scleroderma fibroblasts. Arthritis Rheum 2006 Jul;54(7):2271-2279.

Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics 2004 Sep 1;20(13):2138-2139.

Wei J, Ghosh AK, Sargent JL, Komura K, Wu M, Huang QQ, et al. PPARgamma downregulation by TGFss in fibroblast and impaired expression and function in systemic sclerosis: a novel mechanism for progressive fibrogenesis. PLoS One 2010 Nov 2;5(11):e13778.

Wu M, Melichian DS, Chang E, Warner-Blankenship M, Ghosh AK, Varga J. Rosiglitazone abrogates bleomycin-induced scleroderma and blocks profibrotic responses through peroxisome proliferator-activated receptor-gamma. Am J Pathol 2009 Feb;174(2):519-533.

Xiao R, Kanekura T, Yoshida N, Higashi Y, Yan KL, Fukushige T, et al. 9-Cis-retinoic acid exhibits antifibrotic activity via the induction of cyclooxygenase-2 expression and prostaglandin E2 production in scleroderma fibroblasts. Clin Exp Dermatol 2008 Jul;33(4):484-490.

Xiao R, Yoshida N, Higashi Y, Lu QJ, Fukushige T, Kanzaki T, et al. Retinoic acids exhibit anti-fibrotic activity through the inhibition of 5-lipoxygenase expression in scleroderma fibroblasts. J Dermatol 2011 Apr;38(4):345-353.

Ye Q, Li Y, Jiang H, Xiong J, Xu J, Qin H, et al. Prevention of Pulmonary Fibrosis via Trichostatin A (TSA) in Bleomycin Induced Rats. Sarcoidosis Vasc Diffuse Lung Dis 2014 Oct 20;31(3):219-226.

Zhang XK, Lehmann J, Hoffmann B, Dawson MI, Cameron J, Graupner G, et al. Homodimer formation of retinoid X receptor induced by 9-cis retinoic acid. Nature 1992 Aug 13;358(6387):587-591.

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 Dec;60(12):3807-3814.

TABLES

Table 1 Associations between SSc and haplotypes defined by multi-allelic loci

Physical position[1]	32717094	33043703	33124074	33173650	33302306	33719206	Frequency		OR	95% CI		*P*
Locus	*D6S1100i*	*DPB1*	*D6S0512i*	*D6S2731*	*D6S1701*	*D6S1583*	Case	Control		L	U	
Risk Haplotype1 (RH1)	*114*	**09:01*	*231*	*210*	*162*	*171*	0.226	0.069	4.36	3.18	5.99	8.52×10^{-22}
Risk Haplotype2 (RH2)	-	**13:01*	-	*214*	-	-	0.090	0.011	9.31	4.91	17.6	1.66×10^{-15}

Abbrevations: OR, Odds ratio; CI, confidence Intervals; L, lower; U, upper; *P* , probability.

[1]Physical positions of loci on chromosome 6 are based on the reference UCSC (University of California, Santa Cruz) Genome Browser assembly (GRCh37/hg19) .

Table 2 Nonsynonymous SNVs identical to risk haplotypes in exome sequenicng

						Allele frequency[1]							
						Japanese		Caucasian			Prediction of deleterious SNVs[2]		
Haplotype	SNV	Position[3]	Ref[4]	Alt[5]	Gene	JPT	HGVB	CEU	EA	Subsititution	SIFT	PolyPhen 2	FATHMM
RH1	rs2270191	31080320	C	T	*C6orf15* (NM_014070)	0.1250	0.1197	0.0000	0.0074	exon1:c.G13A:p.V5M	Tolerated	Benign	Tolerated
	rs117951780	31084034	C	T	*CDSN* (NM_001264)	0.1250	0.1197	0.0000	0.0086	exon2:c.G1358A:p.S453N	Tolerated	Possibly damaging	Tolerated
	rs1063646	31107648	C	T	*PSORS1C1* (NM_014068)	0.1394	0.1228	0.1667	0.1659	exon6:c.C398T:p.P133L	Tolerated	Benign	Tolerated
	rs28362675	32362521	C	A	*BTNL2* (NM_019602)	0.2260	0.2190	0.0000	0.0081	exon6:c.G1360T:p.G454C	Tolerated	Benign	Tolerated
	rs41441651	32363888	C	T	*BTNL2* (NM_019602)	0.2260	0.2228	0.0000	0.0090	exon5:c.G1006A:p.D336N	Tolerated	Benign	Tolerated
	rs41417449	32364011	T	C	*BTNL2* (NM_019602)	0.2260	0.2237	0.0000	0.0089	exon5:c.A883G:p.M295V	Tolerated	Benign	Tolerated
	rs34423804	32364046	T	A	*BTNL2* (NM_019602)	0.2260	0.2268	0.0000	0.0089	exon5:c.A848T:p.D283V	Tolerated	Benign	Tolerated
	rs41355746	32364052	C	T	*BTNL2* (NM_019602)	0.2260	0.2208	0.0000	0.0089	exon5:c.G842A:p.R281K	Deleterious	Probably damaging	Tolerated
	rs78587369	32370927	G	A	*BTNL2* (NM_019602)	0.2260	0.2196	0.0000	0.0089	exon3:c.C494T:p.T165I	Tolerated	Possibly damaging	Tolerated
	rs79517313	32713044	C	T	*HLA-DQA2* (NM_020056)	0.1250	0.1170	0.0000	0.0081	exon2:c.C191T:p.T64M	Deleterious	Probably damaging	Tolerated
RH2	rs1042117	33048539	T	A	*HLA-DPB1* (NM_002121)	0.0048	0.0213	0.1111	0.0958	exon2:c.T191A:p.F64Y	Tolerated	Benign	Tolerated
	rs17847931	33167145	A	G	*RXRB* (NM_021976)	0.0048	0.0187	0.0000	0.0001	exon2:c.T284C:p.V95A	Tolerated	Benign	Deleterious

Abbrevations: JPT, Japanese in Tokyo (1000 Genome Brouwser; http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes); HGVB, Human Genetic Variation Database (http://www.genome.med.kyoto-u.ac.jp/SnpDB/); CEU, Utah Residents (1000 Genome Brouwser); EA, European American (NHLBI Exome Sequencing Project: http://evs.gs.washington.edu/EVS/).

[1]Known allele frequency of public database in each population.

[2]Deleterious SNV prediction using multiple method. See Methods.

[3]Physical positions of loci on chromosome 6 are based on the reference UCSC (University of California, Santa Cruz) Genome Browser assembly (GRCh37/hg19) .

[4]Allele of the reference sequence.

[5]Allele identical to risk haplotype.

Table 3 Haplotype associations with SSc by candidate SNVs and multi-allelic loci

Physical position[1]	31084034	32717094	33043703	33124074	33167145	33173650	33302306	33719206	Frequency		OR	95% CI		*P*
Locus	*rs117951780*	*D6S1100i*	*DPB1*	*D6S0512i*	*rs17847931*	*D6S2731*	*D6S1701*	*D6S1583*	Case	Control		L	U	
Risk Haplotype1 (RH1)	A	*114*	**09:01*	*231*	*T*	*210*	*162*	*171*	0.206	0.064	4.10	2.96	5.67	4.63×10^{-19}
	G	*114*	**09:01*	*231*	*T*	*210*	*162*	*171*	0.024	0.005	4.49	1.88	10.7	7.19×10^{-04}
	A or G	*114*	**09:01*	*231*	*T*	*210*	*162*	*171*	0.253	0.079	3.93	3.01	5.14	1.16×10^{-23}
Risk Haplotype2 (RH2)			**13:01*		*C*	*214*			0.088	0.011	9.44	4.97	17.9	1.30×10^{-15}
			**13:01*		*T*	*not 214*			0.006	0.003	1.53	0.38	6.22	1.98×10^{-01}

Abbrevations: OR, Odds ratio; CI, confidence Intervals; L, lower; U, upper; *P* , probability

[1]Physical positions of loci on chromosome 6 are based on the reference UCSC (University of California, Santa Cruz) Genome Browser assembly (GRCh37/hg19) .

Table 4 Association between the number of risk factors and SSc

Number of risk factors	Number		Frequency		Odds ratio	95% CI		*P*
	Case (N=318)	Control (N=561)	Case	Control		L	U	
1	138	85	0.434	0.152	4.29	3.12	5.91	7.55×10^{-20}
1 or 2	169	87	0.531	0.155	6.16	4.44	8.60	3.83×10^{-32}
2	31	2	0.097	0.004	30.2	7.17	127	6.73×10^{-13}

Abbrevations: H, heterozygosity; OR, Odds ratio; CI, confidence interval; L, Lower; U, Upper; *P*, Probability

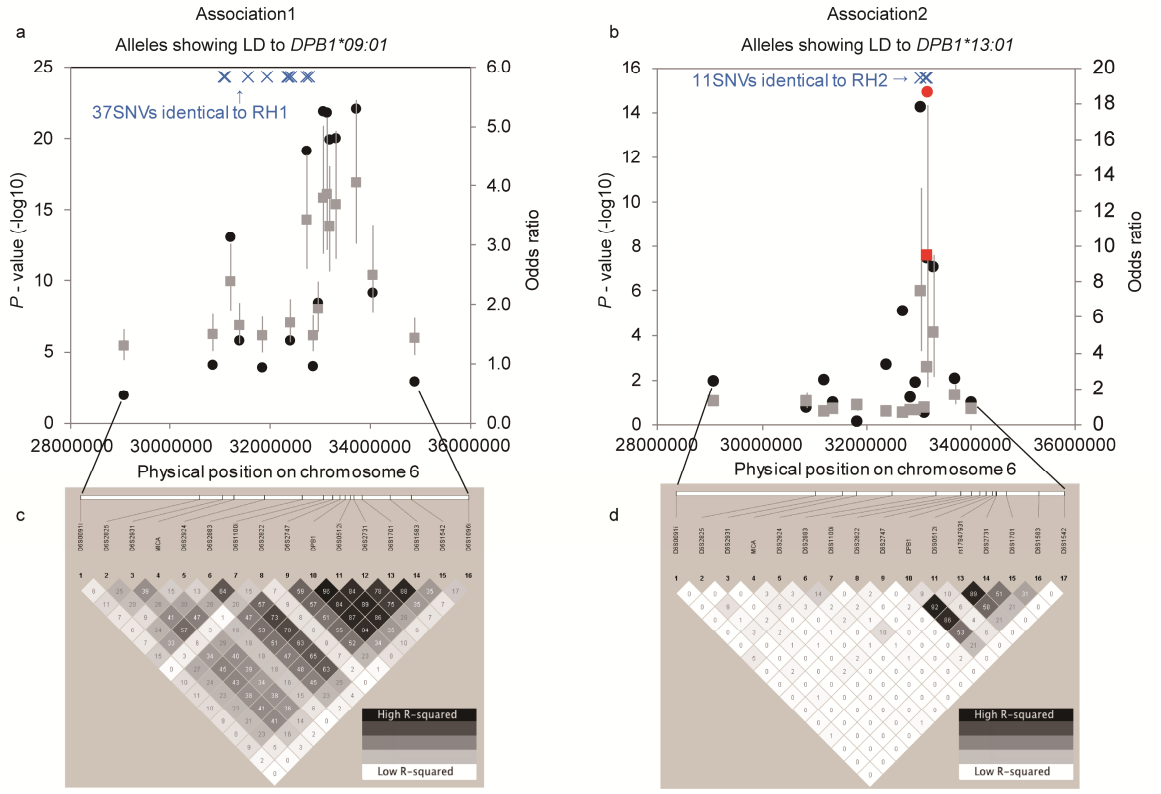
FIGURE LEGENDS

Figure 1. Allelic association and haplotype sequencing within MHC region.

Top figures (a, b) show association peaks for multi-allelic loci and positions of SNV loci with identical risk haplotypes. Black circles show *P*-values for allelic associations (left axis). Gray squares show odds ratios. Gray vertical lines show 95% confidence intervals for odds ratios (right axis). Red circle and square indicate *allele C* of *rs17847931*. Blue crosses show locations of SNVs identical to each risk haplotype (see Supplementary Figure S2, 3 online). Physical positions are based on the reference sequence. Bottom figures (c, d) show pairwise LD (r^2) between alleles of multi-allelic loci.

Figure 2. Haplotype structure harboring allele C of rs17847931.

All haplotype structures harboring alleles for RH2 associated with SSc were found in 3 loci (*DPB1*, *rs17847931*, *D6S2731*) in all subjects. The same haplotypes are abbreviated as dots. Red alleles are those that showed a significant association with SSc in each locus and blue alleles represent the others. Red chromosome IDs indicate homozygotes at *DPB1*13:01* and *allele C* of *rs17847931*. Only 1 haplotype had chromosomes containing *allele C* of *rs17847931*, suggesting no recombinants for the haplotype with the allele within the 3 loci**.** Haplotype and locus association indicated associations between haplotype group and SSc, and between risk allele and SSc in each locus in all subjects. OR, odds ratio; CI, confidence interval; *P*, probability.



Haplotype group	Casse/Control (No. of chromosome)	Chromosome ID	Haplotype			Haplotype association
			DPB1	*rs17847931*	*D6S2731*	
*not *13:01 - T - 214*	Control (25)	IT069_B	*not *13:01*	*T*	*214*	OR = 0.43 95% CI : 0.18 - 1.05 P = 9.27 x 10^{-2}
		▪	▪	▪	▪	
		▪	▪	▪	▪	
		T462_B	*not *13:01*	*T*	*214*	
	Case (5)	TD1032_B	*not *13:01*	*T*	*214*	
		▪	▪	▪	▪	
		▪	▪	▪	▪	
		T0059_A	*not *13:01*	*T*	*214*	
	Case (1)	KA1001_A	**13:01*	*T*	*214*	
**13:01 - C - 214*	Control (12)	IT116_B	**13:01*	*C*	*214*	OR = 9.44 95% CI : 4.97 - 17.9 P = 1.30 x 10^{-15}
		▪	▪	▪	▪	
		▪	▪	▪	▪	
		T455_B	**13:01*	*C*	*214*	
	Case (56)	NS1009_A	**13:01*	*C*	*214*	
		NS1009_B	**13:01*	*C*	*214*	
		FS1001_B	**13:01*	*C*	*214*	
		▪	▪	▪	▪	
		▪	▪	▪	▪	
		TWR0063_B	**13:01*	*C*	*214*	
**13:01 - T - not 214*	Control (3)	T063_A	**13:01*	*T*	*not 214*	OR = 1.53 95% CI : 0.38 - 6.22 P = 1.98 x 10^{-1}
		T063_B	**13:01*	*T*	*not 214*	
		IT250_B	**13:01*	*T*	*not 214*	
	Case (3)	TB1012_A	**13:01*	*T*	*not 214*	
		▪	▪	▪	▪	
		KG1026_B	**13:01*	*T*	*not 214*	
	Allelic association	Allele	**13:01*	*C*	*214*	
		OR	7.40	9.44	3.20	
		95% CI	4.13 - 13.2	4.97 - 17.9	2.09 - 4.91	
		P	6.80 x 10^{-15}	1.30 x 10^{-15}	3.76 x 10^{-08}	