Table 1. Characteristics of the study population*

	Patients	Controls
First set		
Institutions	Kyoto University, Tokyo Women's Medical University	Kyoto University, Tokyo Women's Medical University, BioBank Japan
Typing	TaqMan assay	Illumina HumanHap610 Quad BeadChip, Illumina HumanHap550 BeadChip, Affymetrix Genome-Wide Human SNP Array 6.0
Limited SSc/diffuse SSc, %	49.6/50.4	Not applicable
Anti–topo I/ACA, %	30.6/31.1	Not applicable
Interstitial lung disease, %	48.9	Not applicable
Age, mean ± SD years	50.9 ± 14.7	60.9 ± 12.5
Female, %	91.3	44.9
Replication set		
Institutions	Keio University, Sagamihara National Hospital, Kanazawa University	Kyoto University, BioBank Japan
Typing	TaqMan assay	Illumina HumanHap550 BeadChip, Illumina HumanHap610 Quad BeadChip
Limited SSc/diffuse SSc, %	63.8/34.6	Not applicable
Anti–topo I/ACA, %	29.5/35.2	Not applicable
Interstitial lung disease, %	43.2	Not applicable
Age, mean ± SD years	51.4 ± 14.1	59.3 ± 14.2
Female, %	87.3	48.4

* The first set included 415 patients with systemic sclerosis (SSc) and 16,891 control subjects. The replication set included 315 patients with SSc and 21,054 control subjects. Anti–topo I = anti–topoisomerase I; ACA = anticentromere antibody.

the control subjects with those of patients in the SSc subgroups based on the disease phenotypes. The subanalyses used the same control subjects as were used in the association studies. Intracase analyses based on phenotypes were also performed.

Odds ratios (ORs) and 95% confidence intervals were also calculated. The associations detected in the first and replication studies were then meta-analyzed using the inverse variance method. The resultant *P* values were corrected using the Benjamini-Hochberg false discovery rate (FDR) criterion, and corrected *P* values less than 0.05 were regarded as significant in both the combined study and the subanalyses. The efficiency of the current study was estimated by calculating the likelihood of detecting 3 significant markers (after correcting the *P* values using the FDR method) among 18 randomly selected markers. After the statistically significant markers were identified, the best-fit model for each association was analyzed using dominant, recessive, trend, and allelic chi-square tests or models. Statistical analyses were performed using R or SPSS (version 18) software.

RESULTS

Analyses of candidate genes for SSc in a Japanese population. The 415 patients with SSc and 16,891 control subjects in the first set were genotyped for the 18 markers that were shown to have associations or suspected associations with RA in our previous study. The HLA region was excluded from the genotyped markers, because this region has already been shown to be associated with SSc in Asians. The allele frequencies of the patients were compared with those of the control subjects, using a Cochran-Armitage trend test.

As a result, 3 markers that demonstrated associations with *P* values less than 0.01 in the first set (Table 2) were identified, namely, rs6932056 in the *TNFAIP3* region ($P = 0.0000038$, OR 1.69), rs10821944 in the *ARID5B* region ($P = 0.0025$, OR 1.25), and rs2841277 in the *PLD4* region ($P = 0.0054$, OR 1.25). Two loci that showed suggestive associations with *P* values less than 0.1 (Table 2) were also identified, namely, rs12529514 in the *CD83* region ($P = 0.083$, OR 1.18) and rs2280381 in the *IRF8* region ($P = 0.095$, OR 1.19). The *TNFAIP3* and *IRF8* regions were previously reported to display associations with SSc and lcSSc, respectively, in European populations (10,18). These 5 markers were selected as candidate susceptibility markers for SSc in Japanese and were subjected to validation.

Next, a replication study consisting of 315 patients with SSc and 21,054 control subjects was performed to validate the associations of the 5 markers with SSc. The patients were genotyped for the 5 markers. The genotypes of the control subjects for the 5 markers, except rs6932056, were extracted from the Illumina Infinium HumanHap610 Quad array, as reported previously (31). The genotypes for rs6932056 were imputed based on genome-scanning data using mach2dat soft-

Table 2. Association studies of Japanese patients with SSc*

				Allele 1 frequency								
				First set			Replication set			Combined study		
SNP	Chr	Gene	Allele 1/2	Controls	Patients	*P*	Controls†	Patients	*P*	*P*, patients vs. controls	OR (95% CI)	*P*, patients without overlapping RA vs. controls
rs766449	1	*PADI4*	T/C	0.40	0.37	0.12	–	–	–	–	–	–
rs11900673	2	*B3GNT2*	T/C	0.29	0.28	0.65	–	–	–	–	–	–
rs2867461	4	*ANXA3*	A/G	0.44	0.43	0.57	–	–	–	–	–	–
rs657075	5	*IL3-CSF2*	A/G	0.36	0.34	0.25	–	–	–	–	–	–
rs12529514	6	*CD83*	C/T	0.14	0.16	0.083	0.15	0.16	0.31	0.046	1.15 (1.00–1.33)	0.040
rs1571878	6	*CCR6*	C/T	0.49	0.47	0.28	–	–	–	–	–	–
rs6932056	6	*TNFAIP3*	C/T	0.069	0.11	3.8×10^{-6}	0.067	0.079	0.23	9.5×10^{-6}	1.50 (1.25–1.80)	5.4×10^{-6}
rs2233434	6	*NFKBIE*	G/A	0.21	0.21	0.93	–	–	–	–	–	–
rs10821944	10	*ARID5B*	G/T	0.36	0.41	0.0025	0.36	0.37	0.64	0.0073	1.16 (1.04–1.29)	0.010
rs3781913	11	*PDE2A-CENTD2*	T/G	0.69	0.69	0.91	–	–	–	–	–	–
rs4937362	11	*ETS1-FLI1*	T/C	0.68	0.68	0.88	–	–	–	–	–	–
rs2841277	14	*PLD4*	T/C	0.69	0.74	0.0054	0.69	0.73	0.012	0.00017	1.25 (1.11–1.41)	0.00052
rs3783637	14	*GCH1*	C/T	0.74	0.73	0.54	–	–	–	–	–	–
rs1957895	14	*PRKCH*	G/T	0.39	0.41	0.26	–	–	–	–	–	–
rs6496667	15	*ZNF774*	A/C	0.35	0.37	0.33	–	–	–	–	–	–
rs7404928	16	*PRKCB1*	T/C	0.62	0.63	0.51	–	–	–	–	–	–
rs2280381	16	*IRF8*	T/C	0.84	0.86	0.095	0.83	0.87	0.0099	0.0030	1.26 (1.08–1.47)	0.0021
rs2847297	18	*PTPN2*	G/A	0.34	0.34	0.85	–	–	–	–	–	–

* SSc = systemic sclerosis; SNP = single-nucleotide polymorphism; Chr = chromosome; OR = odds ratio; 95% CI = 95% confidence interval; RA = rheumatoid arthritis.
† The control rs6932056 genotypes used in the replication study were imputed using genome-scanning data obtained for 3,765 subjects.

ware, because rs6932056 was not included in the array. As a result, rs2841277 in the *PLD4* region and rs2280381 in the *IRF8* region showed relatively strong associations with SSc ($P = 0.012$, OR 1.25 and $P = 0.0099$, OR 1.37, respectively) (Table 2). Interestingly, we observed that all 5 of the markers that displayed associations in the first study also demonstrated the same association directions in the replication study.

The inverse variance method was used to combine the data for the first and replication studies. SNPs rs2841277 in the *PLD4* region, rs6932056 in the *TNFAIP3* region, and rs2280381 in the *IRF8* region showed significant associations with SSc even after correcting the associated *P* values using the FDR method for multiple testing (Table 2). Importantly, all 3 of these loci shared risk alleles with RA. Although rs6932056 in the *TNFAIP3* region did not show a strong association with SSc in the replication study, its association was significant in the combined study. The *PLD4* region was shown to be a novel susceptibility gene for SSc, and, for the first time, the *TNFAIP3* and *IRF8* regions were confirmed to be associated with SSc in Japanese.

The association between rs2841277 and SSc was then investigated in detail. When the 200-kbp region around rs2841277 was evaluated, 2 hypothetical genes and cell division cycle associated 4 gene (*CDCA4*) were located at the region, in addition to *PLD4*. *PLD4* was the only gene whose region showed moderate to strong linkage disequilibrium (LD) with rs2841277, indicating *PLD4* as a susceptibility gene (Figure 1A). We vigorously searched candidate markers in exons of *PLD4* that showed strong LD with rs2841277 and selected 2 markers registered in the 1000 Genomes Project (34) that displayed >5% frequency in genotyped subjects, namely, rs2841280 (Figure 1B) and rs894037 in exon 2. Genotyping of these polymorphisms revealed strong LD between rs2841280 (E27Q) and rs2841277 ($D' = 0.98$, $r^2 = 0.75$) and monomorphism of rs894037 in Japanese. An association study of rs2841280 using control genotypes obtained by imputation supported association of *PLD4* with SSc ($P = 6.3 \times 10^{-5}$) (see Supplementary Tables 1 and 2, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.002/art.37777/abstract).

Because the 3 loci were associated with RA in a Japanese population, we analyzed whether the associations with SSc in the current study were contributed by patients with both RA and SSc. When 22 patients who had RA as well as SSc were excluded, significant associations for the 3 loci were still observed (Table 2). A

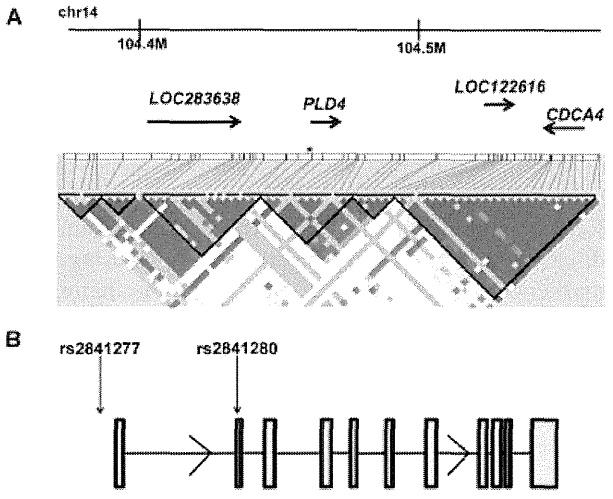

Figure 1. Linkage disequilibrium (LD) block around the *PLD4* region and the *PLD4* structure. **A,** LD block and genes around *PLD4*. The LD block is based on HapMap phase 3 data. **Asterisk** indicates rs2841277. **B,** Schematic view of *PLD4* structure. Rectangles represent exons of *PLD4*.

further stringent analysis excluding patients with other autoimmune diseases demonstrated significant associations of the 3 genes (see Supplementary Table 2). When we compared SSc patients with and those without other autoimmune diseases for the associated alleles, no differences were observed (data not shown).

Subanalysis of types of SSc. Previous studies have revealed that the genetic background of SSc varies between different types of SSc (11,18). Thus, subanalyses of the 5 regions examined in the combined study were performed, in which the allele frequencies of the control subjects were compared with those of the patients with lcSSc or dcSSc. The control subjects were the same as those used in the first study or the combined study. Although *PLD4* and *TNFAIP3* did not display a preference for either SSc phenotype, *IRF8* and *ARID5* showed suggestive preferences for lcSSc, and *CD83* showed a suggestive preference for dcSSc (Table 3).

We also investigated whether the susceptibility loci affect autoantibody status and severe complications. The association studies revealed an association of *TNFAIP3* with SSc patients who possess anticentromere antibodies (ACAs) (see Supplementary Table 3, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37777/abstract), but intracase analyses did not demonstrate clear significance (P = 0.043). We did not observe other associations between the susceptibility loci and clinical phenotypes of SSc, in either case–control analyses or intracase analyses.

Efficacy of the current study. In the current study, a candidate gene analysis was performed based on a meta-analysis of RA GWAS, because many susceptibility genes for autoimmune disease have been reported

Table 3. Associations of the 2 SSc subtypes*

					Limited cutaneous SSc (n = 408)			Diffuse cutaneous SSc (n = 318)		
SNP	Chr	Gene	Allele 1/2	Controls, allele 1 frequency	Allele 1 frequency	P	OR (95% CI)	Allele 1 frequency	P	OR (95% CI)
rs766449	1	*PADI4*	T/C	0.40	0.39	0.52	0.94 (0.77–1.14)	0.36	0.11	0.85 (0.69–1.04)
rs11900673	2	*B3GNT2*	T/C	0.29	0.25	0.096	0.82 (0.66–1.03)	0.31	0.32	1.11 (0.9–1.38)
rs2867461	4	*ANXA3*	A/G	0.44	0.42	0.40	0.92 (0.75–1.12)	0.44	0.97	1.00 (0.82–1.22)
rs657075	5	*IL3-CSF2*	A/G	0.36	0.34	0.54	0.94 (0.76–1.15)	0.33	0.23	0.88 (0.72–1.08)
rs12529514	6	*CD83*	C/T	0.14	0.15	0.79	1.03 (0.85–1.25)	0.18	0.0075	1.32 (1.08–1.62)
rs1571878	6	*CCR6*	C/T	0.49	0.48	0.81	0.98 (0.80–1.19)	0.46	0.20	0.88 (0.72–1.07)
rs6932056	6	*TNFAIP3*	C/T	0.069	0.093	0.0062	1.40 (1.1–1.78)	0.10	0.00063	1.57 (1.21–2.04)
rs2233434	6	*NFKBIE*	G/A	0.21	0.20	0.60	0.94 (0.73–1.20)	0.22	0.70	1.05 (0.83–1.33)
rs10821944	10	*ARID5B*	G/T	0.36	0.40	0.0085	1.22 (1.05–1.41)	0.38	0.30	1.09 (0.93–1.29)
rs3781913	11	*PDE2A-CENTD2*	T/G	0.69	0.69	0.98	1.00 (0.81–1.24)	0.69	0.90	1.01 (0.82–1.25)
rs2841277	14	*PLD4*	T/C	0.69	0.73	0.0067	1.24 (1.06–1.45)	0.74	0.0049	1.29 (1.08–1.55)
rs2841280	14	*PLD4*	C/G	0.64	0.69	0.0011	1.30 (1.11–1.52)	0.69	0.0086	1.27 (1.06–1.51)
rs2847297	18	*PTPN2*	G/A	0.34	0.33	0.67	0.96 (0.78–1.18)	0.34	0.87	1.02 (0.83–1.25)
rs4937362	11	*ETS1-FLI1*	T/C	0.68	0.68	0.75	0.97 (0.78–1.19)	0.69	0.92	1.01 (0.82–1.25)
rs3783637	14	*GCH1*	C/T	0.74	0.73	0.69	0.96 (0.77–1.19)	0.73	0.65	0.95 (0.76–1.18)
rs1957895	14	*PRKCH*	G/T	0.39	0.40	0.84	1.02 (0.84–1.25)	0.42	0.16	1.15 (0.95–1.41)
rs6496667	15	*ZNF774*	A/C	0.35	0.39	0.088	1.19 (0.97–1.45)	0.34	0.75	0.97 (0.79–1.19)
rs7404928	16	*PRKCB1*	T/C	0.62	0.61	0.60	0.95 (0.78–1.16)	0.66	0.15	1.17 (0.95–1.44)
rs2280381	16	*IRF8*	T/C	0.84	0.88	0.0038	1.36 (1.11–1.68)	0.86	0.21	1.16 (0.92–1.45)

* SSc = systemic sclerosis; SNP = single-nucleotide polymorphism; Chr = chromosome; OR = odds ratio; 95% CI = 95% confidence interval.

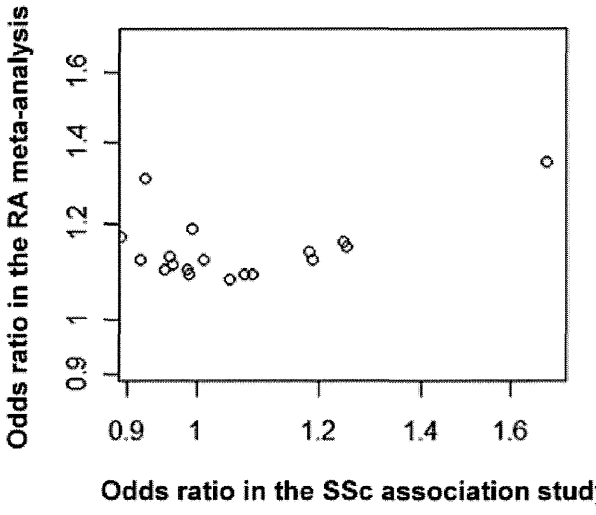

Figure 2. Comparison of associations for systemic sclerosis (SSc) and rheumatoid arthritis (RA). The odds ratios obtained for 18 genes in association studies of SSc and RA are plotted.

to be shared by a wide range of diseases. As a result, 3 susceptibility genes for SSc in Japanese were identified. Thus, we analyzed whether the candidate gene approach taken in the current study for detecting novel susceptibility genes for SSc was effective. When the likelihood of finding 3 susceptibility genes among 18 genes by chance was calculated, the likelihood was determined to be 2.5×10^{-8}. These results indicated that our approach to identifying novel susceptibility genes for systemic diseases is effective. It would be interesting to compare the risk direction of the genotyped markers between RA and SSc. Although the 3 susceptibility loci for SSc shared risk direction with RA, no correspondence of the risk directions of the markers between the 2 diseases was detected (Figure 2). This indicated that a large proportion of the 18 RA markers are not shared by SSc, and that the lack of association between the 13 markers and SSc was not attributable to the low power produced by the relatively small number of SSc patients included in this study.

DISCUSSION

Because SSc can lead to severe complications, poor quality of life, and shortened survival, clarifying the characteristics of SSc is important. Clarification of the disease would aid the search for novel therapeutic targets and the development of new therapeutic strategies. Detecting susceptibility genes using GWAS or a candidate gene approach would also help to uncover the pathophysiology underlying SSc.

Previous studies have revealed that more than 15 markers and loci are associated with SSc. However, the markers detected so far cannot fully explain the genetics of SSc, indicating that many susceptibility genes are yet to be identified. Because a relatively large proportion of RA susceptibility genes are shared by other autoimmune diseases (24), a candidate gene approach using novel markers observed in GWAS of RA is a fascinating way of identifying new SSc markers. In fact, some of the novel susceptibility markers for RA identified in the meta-analysis were shown to be susceptibility markers for systemic lupus erythematosus (SLE) and Graves' disease (31).

In the current study, we successfully identified 3 susceptibility genes for SSc in Japanese. No studies have identified *PLD4* as an SSc-associated locus. The current study is also the first to detect *TNFAIP3* and *IRF8* as susceptibility genes for SSc in a Japanese population. The best-fit models for each association are shown in Supplementary Table 4, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.002/art.37777/abstract.

It is conceivable that these 3 associations might have been obtained due to the overlap of RA and SSc. Even after excluding the patients with both RA and SSc based on physicians' reports, the significant associations for the 3 loci were still observed (Table 3). Information regarding rheumatoid factor (RF) and anti–citrullinated protein antibody (ACPA) was available for 371 SSc patients without RA and 65 SSc patients without RA, respectively, of whom 21.6% and 10.8% were positive for RF and ACPA, respectively. These prevalences are compatible with those previously observed in SSc patients without RA (35,36). Moreover, we showed that the effect sizes and risk direction of the markers tested in this study were dissociated between SSc and RA. In addition, further stringent analysis comprising SSc patients without any autoimmune disease also showed the associations of the 3 loci. These results indicate that the associations of the 3 loci are not attributable to overlapping of RA or other diseases.

Although the associations of the *ARID5B* and *CD83* loci with SSc did not reach a stringently significant level in the combined study, the tendencies toward an association with SSc displayed by rs10821944 in the *ARID5B* locus and rs12529514 in the *CD83* region in the first study were maintained in the replication study. This indicates that these loci are potential susceptibility regions for SSc. Further replication studies are needed to

address the associations of these 2 loci with SSc in a Japanese population.

Because *TNFAIP3* was reported to be strongly associated with SSc in a European population (18), the significant associations detected in the combined study indicate that *TNFAIP3* displays general associations with SSc that go beyond ethnic boundaries. In addition, rs6932056, which displayed a strong association with SSc in a European population (18), is in strong LD with rs5029939 ($r^2 = 0.85$) in the Japanese population. SNP rs6932056 also displays strong LD with rs2230926, a missense mutation of *TNFAIP3* ($r^2 = 0.85$), in Japanese. The rs2230926 missense mutation leads to an amino acid alteration in the OTU (ovarian tumor) domain of the A20 protein, which is considered to result in decreased NF-κB signaling. Because we did not observe strong associations between rs6932056 and SSc in the replication study, it will be necessary to reexamine the association between *TNFAIP3* and SSc using independent sample sets of Japanese patients with SSc, in spite of the significant associations detected in this study.

PLD4 is a recently reported member of the phospholipase family without phospholipase D activity. *PLD4* is expressed in the spleen and early postnatal microglia in the white matter of mice (37). The phenotypes of *Pld4*-deficient mice have not been reported. In addition, little is known about the expression or distribution of *PLD4* in humans. Although the functions of *PLD4* are also poorly understood, it is known to be involved in the phagocytosis of microglia (38). The expression of *PLD4* around the marginal zone in the spleen might support the functional involvement of *PLD4* in immunologic systems. It is interesting that rs2841280, which alters an amino acid of PLD-4, is associated with SSc. Minor allele G of rs2841280 is associated in a protective manner. The impact of an amino acid alteration brought by rs2841280 on the effect of PLD-4 protein is not known.

When we analyzed the impact of the amino acid alteration using in silico analysis (SIFT software; http://sift.jcvi.org/), it was shown to result in a small effect. However, the association raises the possibility that this polymorphism leads functional modulation of PLD-4, and it is feasible to analyze the functional change of PLD-4 protein with rs2841280, using animal models of SSc. When we performed an in silico analysis of the effect of rs2841277 and rs2841280 on *PLD4* expression, we did not detect any clear associations between the 2 genotypes and *PLD4* transcription ($P > 0.05$) (39). Therefore, in spite of the association of these 2 mutations, it has not been confirmed whether one of these 2 polymorphisms is the causative mutation.

Although the detection of a P value less than 5×10^{-8} in a GWAS is stringent evidence of an association between a marker and a particular disease, the detection of suggestive associations between the *PLD4* region and SSc in European GWAS would indicate that associations exist between *PLD4* and SSc in other populations. However, when we examined the associations between the *PLD4* locus or nearby loci and SSc in GWAS involving a European population, we did not detect any strong associations ($P < 10^{-4}$) (8,9). According to the HapMap database, the European population displays a higher risk allele frequency for rs2841277 than the Japanese population. In addition, the HapMap database also indicates that the LD block spanning *PLD4*, which includes rs2841277, is similar in Europeans and Japanese. Nevertheless, a European population did not show a strong association between *PLD4* and SSc, suggesting that *PLD4* has a stronger effect on autoimmune diseases in Japanese than in Europeans. There is also a possibility that these 2 polymorphisms are only markers, and that a rare variant in LD with the 2 markers affects disease onset. A rare causative variant might explain a different association of *PLD4* with SSc between populations.

IRF8 was shown to be associated with SLE in a European population (40). Interferon regulatory factor 8 (IRF-8) protein is a transcription factor involved in the interferon pathway. The interferon pathway has been shown to be involved with a broad range of autoimmune diseases, including SSc (41). Thus, it is interesting that *IRF5* and *IRF8*, both of which belong to the IRF family, displayed associations with SSc. Although a European GWAS of SSc patients revealed suggestive associations between the *IRF4* locus and SSc, the results were not successfully replicated (8), indicating that the different functional roles of each IRF family molecule might influence the development of SSc. *IRF8* promotes B cell differentiation; however, the roles and importance of B cells in skin fibrosis in SSc patients have not been established (42–44). *IRF8* and its mutant variants are also known to be involved in the development of dendritic cells (45). Thus, the association between *IRF8* and SSc might indicate the involvement of B cells and dendritic cells in the development of SSc.

When the patients with SSc were classified as having either lcSSc or dcSSc and subanalyses were performed, *ARID5B*, *IRF8*, and *CD83* displayed stronger associations with one of the 2 phenotypes. However, the associations of these 3 markers with the phenotypes

were not strong enough to provide convincing evidence of a clear distinction between the genetic backgrounds of the 2 SSc phenotypes. When the associations of the SSc subtypes with the other 13 markers in the first set were analyzed, no strong association was detected ($P > 0.05$). Other subanalyses of the susceptibility loci in the combined set did not show significant results between disease phenotypes, due to lack of power. Because classification according to disease phenotypes resulted in limited numbers of subjects in each subset, we conducted this subanalysis only in the combined set. The association between *TNFAIP3* and ACAs should be confirmed in a large-scale association study.

Although GWAS are an extremely powerful way to detect novel susceptibility genes for diseases, GWAS of patients with SSc have been performed only in European populations. Our study detected strong evidence for the sharing of susceptibility genes between RA and SSc in a Japanese population. In addition, the current study indicated that a candidate gene approach based on the results of GWAS of other diseases that display pathologic signaling pathways or mechanisms similar to those associated with the disease being examined is an effective approach to identifying novel susceptibility genes.

It will be interesting to perform GWAS of Japanese patients with SSc and analyze the similarities and differences in the detected associations not only between Japanese and Europeans but also between Japanese patients with SSc and Japanese patients with RA.

ACKNOWLEDGMENT

We thank the staff of the BioBank Japan Project for collecting DNA samples from control subjects.

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. Terao 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. Terao, Ohmura, Kawaguchi, Nishimoto, Kawasaki, Takehara, Furukawa, Kochi, Ota, Ikari, Sato, Tohma, Yamada, Yamamoto, Kubo, Yamanaka, Kuwana, Tsuchiya, Matsuda, Mimori.

Acquisition of data. Terao, Ohmura, Kawaguchi, Nishimoto, Kawasaki, Takehara, Furukawa, Kochi, Ota, Ikari, Sato, Tohma, Yamada, Yamamoto, Kubo, Yamanaka, Kuwana, Tsuchiya, Matsuda, Mimori.

Analysis and interpretation of data. Terao, Ohmura.

REFERENCES

1. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum 2008;37:223–35.
2. Kawut SM, Taichman DB, Archer-Chicko CL, Palevsky HI, Kimmel SE. Hemodynamics and survival in patients with pulmonary arterial hypertension related to systemic sclerosis. Chest 2003;123:344–50.
3. 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.
4. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. Nat Rev Rheumatol 2012;8:42–54.
5. Romano E, Manetti M, Guiducci S, Ceccarelli C, Allanore Y, Matucci-Cerinic M. The genetics of systemic sclerosis: an update. Clin Exp Rheumatol 2011;29:S75–86.
6. 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.
7. Terao C, Ohmura K, Katayama M, Takahashi M, Kokubo M, Diop G, et al. Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis: a genome-wide study combined with immunological analyses. PLoS One 2011;6:e20457.
8. 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.
9. 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;7:e1002091.
10. 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;7:e1002178.
11. Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suarez H, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet 2009;18:2071–7.
12. Dieude P, Guedj M, Wipff J, Avouac J, Fajardy I, Diot E, et al. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. Arthritis Rheum 2009;60:225–33.
13. Gourh P, Agarwal SK, Divecha D, Assassi S, Paz G, Arora-Singh RK, et al. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis: evidence of possible gene–gene interaction and alterations in Th1/Th2 cytokines. Arthritis Rheum 2009;60:3794–806.
14. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Airo P, et al. NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis. Ann Rheum Dis 2011;70:668–74.
15. Bossini-Castillo L, Broen JC, Simeon CP, Beretta L, Vonk MC, Ortego-Centeno N, et al. A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. Ann Rheum Dis 2011;70:638–41.
16. Dieude P, Guedj M, Truchetet ME, Wipff J, Revillod L, Riemekasten G, et al. Association of the CD226 Ser^{307} variant with systemic sclerosis: evidence of a contribution of costimulation pathways in systemic sclerosis pathogenesis. Arthritis Rheum 2011;63:1097–105.
17. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. J Autoimmun 2010;34:155–62.
18. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M, et al. Association of the TNFAIP3 rs5029939 variant

with systemic sclerosis in the European Caucasian population. Ann Rheum Dis 2010;69:1958–64.
19. Tsuchiya N, Kawasaki A, Hasegawa M, Fujimoto M, Takehara K, Kawaguchi Y, et al. Association of STAT4 polymorphism with systemic sclerosis in a Japanese population. Ann Rheum Dis 2009;68:1375–6.
20. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Hikami K, et al. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum 2009;60:1845–50.
21. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Kawamoto M, et al. Association of the FAM167A–BLK region with systemic sclerosis. Arthritis Rheum 2010;62:890–5.
22. Hasebe N, Kawasaki A, Ito I, Kawamoto M, Hasegawa M, Fujimoto M, et al. Association of UBE2L3 polymorphisms with diffuse cutaneous systemic sclerosis in a Japanese population. Ann Rheum Dis 2012;71:1259–60.
23. 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.
24. Suzuki A, Kochi Y, Okada Y, Yamamoto K. Insight from genome-wide association studies in rheumatoid arthritis and multiple sclerosis. FEBS Lett 2011;585:3627–32.
25. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis: a genomewide study. N Engl J Med 2007;357:1199–209.
26. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977–86.
27. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–82.
28. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nat Genet 2007;39: 1431–3.
29. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.
30. Diaz-Gallo LM, Gourh P, Broen J, Simeon C, Fonollosa V, Ortego-Centeno N, et al. Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Ann Rheum Dis 2011; 70:454–62.
31. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 2012;44:511–6.
32. 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.
33. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
34. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010;467: 1061–73.
35. Mimura Y, Ihn H, Jinnin M, Asano Y, Yamane K, Yazawa N, et al. Rheumatoid factor isotypes and anti-agalactosyl IgG antibodies in systemic sclerosis. Br J Dermatol 2004;151:803–8.
36. Santiago M, Baron M, Miyachi K, Fritzler MJ, Abu-Hakima M, Leclercq S, et al. A comparison of the frequency of antibodies to cyclic citrullinated peptides using a third generation anti-CCP assay (CCP3) in systemic sclerosis, primary biliary cirrhosis and rheumatoid arthritis. Clin Rheumatol 2008;27:77–83.
37. Yoshikawa F, Banno Y, Otani Y, Yamaguchi Y, Nagakura-Takagi Y, Morita N, et al. Phospholipase D family member 4, a transmembrane glycoprotein with no phospholipase D activity, expression in spleen and early postnatal microglia. PLoS One 2010;5: e13932.
38. Otani Y, Yamaguchi Y, Sato Y, Furuichi T, Ikenaka K, Kitani H, et al. PLD4 is involved in phagocytosis of microglia: expression and localization changes of PLD4 are correlated with activation state of microglia. PLoS One 2011;6:e27544.
39. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315: 848–53.
40. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet 2011;7:e1002341.
41. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Ann Rheum Dis 2011;70:2029–36.
42. Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A 2003;100:12319–24.
43. Lafyatis R, Kissin E, York M, Farina G, Viger K, Fritzler MJ, et al. B cell depletion with rituximab in patients with diffuse cutaneous systemic sclerosis. Arthritis Rheum 2009;60:578–83.
44. Smith V, Van Praet JT, Vandooren B, Van der Cruyssen B, Naeyaert JM, Decuman S, et al. Rituximab in diffuse cutaneous systemic sclerosis: an open-label clinical and histopathological study. Ann Rheum Dis 2010;69:193–7.
45. Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, et al. IRF8 mutations and human dendritic-cell immunodeficiency. N Engl J Med 2011;365:127–38.

Review Article

Recent advances on the genetics of rheumatoid arthritis: current topics and the future

Koichiro Ohmura[1,*], Chikashi Terao[2] and Tsuneyo Mimori[1]

[1]Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan

[2]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan

Rheumatoid arthritis (RA) is a chronic autoimmune disease that causes severe joint pain and eventually joint deformity. Recent large cohort studies and the rapid progression of genotyping platforms have enabled identification of more than 30 susceptibility genes for RA. *HLA* is the major genetic determinant for RA for which a shared epitope hypothesis (70th-74th amino acids of HLA-DR β chain determine susceptibility) has been accepted. However, recent detailed single nucleotide polymorphism (SNP) typing of the *HLA* region and imputation method revealed that the most important amino acid positions of the HLA-DR β chain are the 11th in addition to the 71st and the 74th. HLA-B (at position 9) and HLA-DPB1 (at position 9) are also important determinants. This revised shared epitope hypothesis will form a new theory for *HLA* association. Another topic is that anti-citrullinated protein antibody (ACPA)-negative RA has been shown to be genetically different from ACPA-positive RA. Many susceptibility genes including *HLA* were not associated with ACPA-negative RA; however, we have shown that some *HLA* alleles are associated with ACPA-negative RA. In this review, we present some new findings regarding *HLA* as well as some recently discovered susceptibility genes for RA.

Rec.4/6/2012, Acc.5/14/2012, pp90-98

*Correspondence should be addressed to:
Koichiro Ohmura, Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. Phone: +81-75-751-4380, Fax: +81-75-751-4338, Email: ohmurako@kuhp.kyoto-u.ac.jp

Key words genetics, GWAS, HLA, rheumatoid arthritis, SNP

Introduction

Since 2003, when sequencing of the human genome was completed, there has been a burst of identification of new susceptibility genes for RA. In the last several years in particular, more than 30 genes or loci have been identified as RA-related genes[1]. This activity was supported by the development of SNP genotyping platform, which enables us to type hundreds of millions of SNPs in a few weeks, even in a relatively small lab. In addition, a growing number of large cohorts were formed to tackle the elucidation of RA pathogenesis, which provided substantial power to detect genes of significance[2].

ACPA is a specific autoantibody of RA, and its target antigens are citrullinated vimentin, filaggrin, α-enolase, and others[3]. It is a useful marker not only for diagnosis of RA, but also for predicting disease course[4]. ACPA-positive RA

is clinically severer than ACPA-negative RA. Moreover, it has been suggested that ACPA-positive RA is genetically distinct from ACPA-negative RA[5,6].

Here we present the recent advances in RA genetics and also discuss the genetic differences between ACPA-positive and ACPA-negative RA.

Human leukocyte antigen (HLA)

Genetic predisposition to RA has been investigated intensively. *HLA* is a major determinant of RA susceptibility and *HLA-DRB1**01:01, *01:02, *04:01, *04:04, *04:05, *04:08, *04:10, *04:13, *04:16, *10:01, *14:02 and *14:06 were reported to be associated with RA development. Among these *HLA-DRB1* alleles, there are common amino acid sequences at the 70th-74th residues of the HLA-DRβ chain (QKRAA, QRRAA or RRRAA), which is called a 'shared epitope' (SE)[7]. The association of *HLA-DRB1* SE with RA has been replicated in many ethnic groups[8]. However, recently the important role of Leucine at 67th position (Leu67)[9-10] and Valine at 11th position (Val11)[10] for RA development and resistant effect on RA development by Aspatic acid at 70th position (Asp70)[11] were also reported. In addition, Raychaudhuri et al. used existing genome-wide SNP data of >5,000 ACPA-positive RA cases and ~15,000 controls and imputed (expected SNP genotypes in silico from adjacent SNP gentotypes and linkage disequilibrium information) the gap SNP genotypes of HLA locus and reported the following findings. They showed that three amino acid positions (11, 71 and 74) of HLA-DRβ chain as well as single-amino acid positions in HLA-B (at position 9) and HLA-DPβ chain (at position 9) explain most of the MHC association with RA[12]. All these positions are located in peptide-binding grooves, as shown in Fig.1. Among these positions, position 11 of HLA-DRβ chain showed the strongest association with RA development ($p<10^{-561}$ for position 11). As shown in Table 1, Val11 and Leu11 are the key amino acids for susceptibility and

Table 1 Effect estimates of the 3 amino acids associated with risk of RA

HLA-DRβ1 amino acid at position			multivariate OR	95%CI	*HLA-DRB1* alleles
11	71	74			
Val	Lys	Ala	4.44	4.02-4.91	*04:01
Val	Arg	Ala	4.22	3.75-4.75	*04:08, *04:05, *04:04, *10:01
Leu	Arg	Ala	2.17	1.94-2.42	*01:02, *01:01
Pro	Arg	Ala	2.04	1.59-2.62	*16:01
Val	Arg	Glu	1.65	1.24-2.19	*04:03, *04:07
Asp	Arg	Glu	1.65	1.29-2.10	*09:01
Val	Glu	Ala	1.43	1.04-1.96	*04:02
Pro	Ala	Ala	1.00	Reference	*15:01, *15:02
Ser	Arg	Ala	0.88	0.77-1.00	*11:01, *11:04, *12:01
Ser	Arg	Leu	0.71	0.57-0.89	*08:01, *08:04
Ser	Lys	Arg	0.63	0.54-0.73	*03:01
Ser	Glu	Ala	0.59	0.51-0.68	*11:02, *11:03, *13:01, *13:02

Estimate effects for haplotypes of *HLA-DRB1*. For each haplotype, the multivariate effect is given as an odds ratio (OR), taking the most frequent haplotype (Pro-Ala-Ala) in the control samples as the reference (that is, given that the haplotype has an OR of 1). Classical shared epitope alleles are shown in bold. This table is modified from a previous report[12].

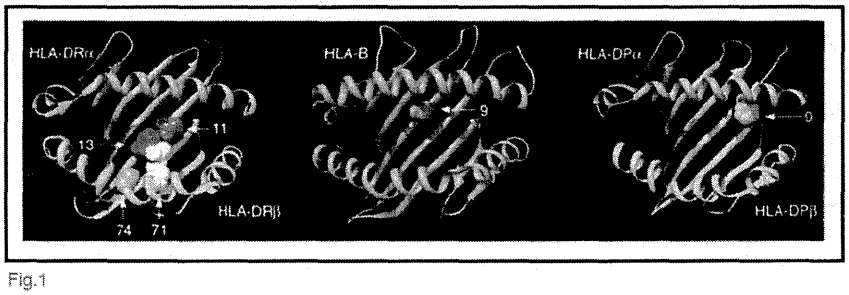

Fig.1
Three-dimensional ribbon models for the HLA-DR, HLA-B and HLA-DP proteins. Key amino acid positions identified by the association analysis are highlighted. This figure is taken from a previous report[12].

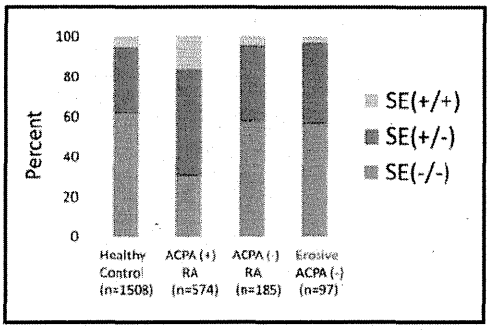

Fig.2
Prevalence of individuals carrying double SE, single SE or no SE is shown in healthy control, ACPA-positive RA, ACPA-negative RA and ACPA-negative RA with typical bone erosion as determined by X-ray. This clearly shows that ACPA-negative RA is distinct from ACPA-positive RA. This figure is illustrated based on our previous report[13].

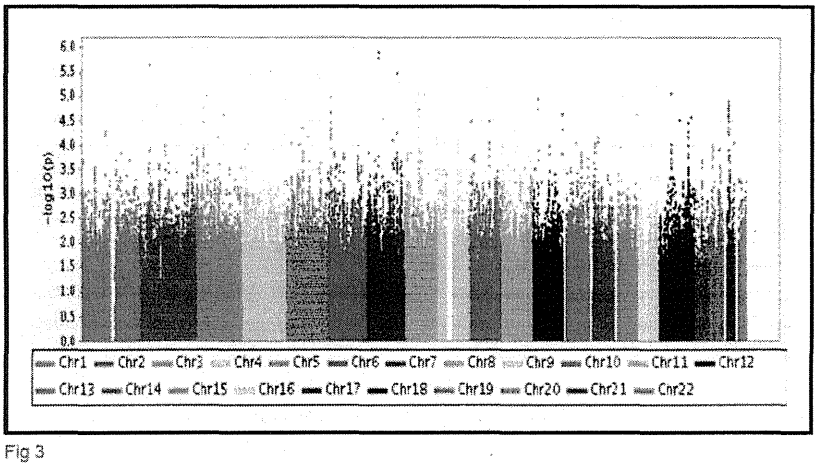

Fig 3
Probability plot for association with ACPA-negative RA (n=774) versus healthy controls (n=1079). This figure is taken from a previous report[6].

Ser11 is protective, for example, even though positions 71 and 74 are the SE types, Ser11 offsets such effects. Since most of the SE alleles have Valine or Leucine at position 11, Leucine at position 67, and do not have Serine at position 11 nor Aspatic acid at position 70, the results of previous studies using SE would not have been affected by the recent findings. Thus, key amino acid positions of HLA-DRβ chain for RA development seem to be 11th, 70th, 71st, and 74th positions and there still are some debates which positions have the primary effect. Anyway, these positions seem to be important for citrullinated peptide presentation.

HLA association with ACPA-negative RA

In 2005, a Dutch group reported that the association of SE was only exhibited with ACPA-positive RA and no association was seen with the ACPA-negative RA patients[10]. We have replicated the results in the Japanese population, and also showed that similar results were obtained even when we selected only bone-erosive ACPA-negative RA[13], which strongly suggests that this observation is not due to the contamination of non-RA arthritic diseases in ACPA-negative RA subset (Fig.2).

First of all, is there a genetic predisposition for ACPA-negative RA? From a twin study, heritability of ACPA-negative RA has been estimated and is thought to be as high as that of ACPA-positive RA[14].

Next, is *HLA* associated with ACPA-negative RA? A genome-wide association study (GWAS) meta-analysis of ACPA-negative RA showed that *HLA-DR* locus in chromosome 6 had no peak of association (see Fig.3)[6], suggest-

ing that the impact of *HLA* for development of ACPA-negative RA is not as large as that of ACPA-positive RA. In the study, the p-value of the *HLA* locus for ACPA-positive RA reached the order of 10^{-60}; in contrast, that for ACPA-negative RA reached the order of 10^{-4}. However, this does not mean that *HLA* is not associated with ACPA-negative RA, but probably means that ACPA-positive RA is a rather homogeneous subset in terms of *HLA* usage compared with ACPA-negative RA. ACPA-negative RA might have more variations of autoantigen (probably not citrullinated). In ACPA-positive RA, *HLA* usage is rather homogeneous, probably because citrullinated proteins or peptides are the common autoantigens among such patients that have SE-carrying *HLA*.

What *HLA* alleles are associated with ACPA-negative RA? In Caucasians, *HLA-DR3* and *DR13* have been reported to be associated with ACPA-negative RA[15-17]. As *HLA-DR3* association was seen in 3 independent European cohorts, it is probably true in Caucasians. In Japanese, we found that multiple *HLA-DRB1* alleles, including *12:01, *14:03 and *04:05, were associated with ACPA-negative RA susceptibility in the Japanese population[18]. *HLA-DR3* alleles were not shown because they are very rare in Japanese. We also found that *HLA-DRB1**15:02 and *13:02 were protective against ACPA-negative RA development. It is noteworthy that one of the SE alleles, *HLA-DRB1**04:05, was associated with ACPA-negative RA. Other SE alleles were not associated with ACPA-negative RA. This implies that the association of *04:05 with ACPA-negative RA is not due to the common amino acid sequence of SE because SE-carrying alleles other than *04:05 are not associated. Therefore, other mechanisms are suggested.

It seems there are two subsets in ACPA-negative RA based on RF positivity. Mackie et al. recently reported that *HLA-DRB1* SE is associated with ACPA(-)RF(+) RA but not with ACPA(-) RF(-) RA[19]. We have similar data for the Japanese population and showed that there are some specific *HLA-DRB1* alleles associated with ACPA(-) RF(+) RA or ACPA(-)RF(-) RA (Fig.4). For example, *04:05 and *09:01 were specifically associated with ACPA(-)RF(+) subset, and DR8/DR8 homozygote and DR14 were specifically associated with ACPA(-)RF(-) subset, whereas *12:01 was associated with both subsets. In contrast, ACPA(+)RA could not be separated by *HLA-DR* allelic usage.

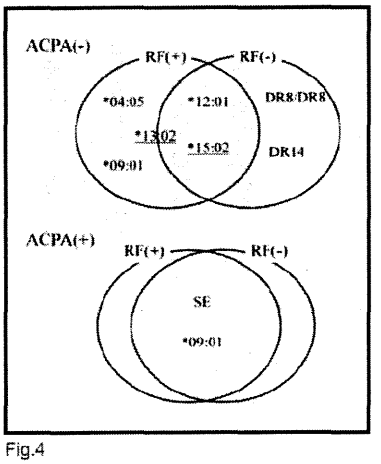

Fig.4
Scheme of HLA-DRB1 allele association with RF(+) or RF(-) subset of ACPA(-) RA or ACPA(+)RA in Japanese. Underline represents the protective allele. This figure is taken from our unpublished results.

Non-HLA genes associated with RA

A lot of genetic polymorphisms of candidate genes were tested for association with RA and reported to be associated with it, but most of them were not replicated. Perhaps the positive results are due to publication bias and relatively small sample sizes. Since 2003[20], genome-wide association studies (GWAS) have been applied to RA[21-26] and recently several meta-analyses of GWAS were performed[27-29]. Sample sizes also jumped from several hundred to tens of thousands. As a result, 30-40 genes or loci were detected to be significantly ($p<5\times10^{-8}$) associated with RA[1]. Many of these SNPs are located not in the genes (exons and introns), but near the genes, while some of the SNPs are located in exons and cause amino acid substitution (e.g. *PTPN22*). In many cases, the real causative SNPs or variants are still unknown. The list of SNPs in Table 2 shows the most strongly associated SNPs in the studies, but the real causative variants may exist somewhere else. The associated genes shown in Table 2 are classified by their main function. These genetic variants satisfied the genome-wide significance ($p<5\times10^{-8}$) or region-wide significance after Bonferroni's correction with multiple replication. Some of them are specific to Caucasians, mainly due to the absence of polymorphisms such as *PTPN22* and *RBPJ*, while some are specific to the Japanese or

Table 2 Candidate genes with confirmed association with rheumatoid arthritis

Gene	Best p-value	OR	Association† in		landmark SNP	SNP position	reference
			Caucasians	Japanese*			
(1)Intracellular signaling molecules and receptors							
PTPN22	9.1×10^{-24}	1.94	++	NA	rs2476601	exon	27
TRAF1-C5	4.0×10^{-14}	1.32	++	-	rs3761847	near	22
MBP	2.7×10^{-6}	1.23	-	++	rs2000811	intron	26
TNFAIP3	8.9×10^{-13}	1.22	++	++	rs6920220	near	27
BLK	5.7×10^{-9}	1.19	++	+	rs2736340	near	24
SPRED2	5.3×10^{-10}	1.13	++	+	rs934734	intron	27
TAGAP	3.8×10^{-7}	0.91	+	-	rs394581	near	44
TRAF6	3.9×10^{-6}	0.89	+	-	rs540386	intron	44
PTPRC	6.7×10^{-7}	0.88	+	-	rs10919563	intron	44
PRKCQ	4.4×10^{-6}	0.88	+	-	rs4750316	near	45
(2)Transcription factor							
REL	3.1×10^{-14}	1.25	++	-	rs13031237	intron	24
IRF5	4.2×10^{-11}	1.25	++	+	rs10488631/ rs13225818	near/near	27
STAT4	1.7×10^{-11}	1.24	++	++	rs7574865	intron	46
RBPJ	1.0×10^{-16}	1.18	++	NA	rs874040	near	27
AIRE	3.6×10^{-6}	1.18	-	++	rs2075876	intron	33
AFF3	1.0×10^{-14}	1.15	++	+	rs11676922	near	27
PRDM1	2.1×10^{-6}	1.11	++	-	rs6822844	near	44
(3)Cytokines and cytokine receptors							
CCR6	7.7×10^{-19}	1.19	++	++	rs3093024	near	25
IL2RB	4.6×10^{-6}	1.13	++	-	rs3218253	intron	47
IL2RA	1.4×10^{-11}	1.11	++	-	rs706778	intron	27
TNFRSF14	1.1×10^{-7}	0.92	+	+	rs3890745	near	45
CCL21	3.9×10^{-10}	0.87	++	-	rs951005	near	27
ANKRD55-IL6ST	9.6×10^{-12}	0.85	++	-	rs6859219	near	27
IL2-IL21	5.6×10^{-5}	0.78	+	NA	rs6822844	near	46
(4)Membrane receptors and costimulatory molecules							
HLA-DRB1	$<10^{-299}$	2.88	++	++	rs6910071	exon	27
FCRL3	8.5×10^{-7}	2.15	+	+	rs10430455	near	48
CD244	7.0×10^{-6}	1.31	-	+	rs6682654	intron	49
CD2-CD58	1.0×10^{-9}	1.13	++	-	rs11586238	near	44
CD28	1.3×10^{-9}	1.13	++	-	rs1980422	near	44
FCGR2A	1.5×10^{-5}	1.12	+	NA	rs12746613	near	44
CTLA4	6.3×10^{-6}	0.86	++	+	rs231735	near	27
CD40	2.8×10^{-9}	0.85	++	-	rs4810485	intron	27
(5)Enzymes							
PADI4	4.6×10^{-6}	1.97	+	++	rs766449	intron	20
PXK	3.1×10^{-14}	1.13	++	NA	rs13315591	near	27
DDX6	1.1×10^{-8}	0.87	++	-	rs10892279	near	28
(6)Unknown							
KIF5A-PIP4K2C	8.8×10^{-6}	0.89	+	-	rs1678542	near	45
C5orf30	4.1×10^{-6}	0.93	++	-	rs26232	intron	27

NA: not applicable due to the lack of polymorphism in Japanese

*Associations in Japanese are mainly based on our recent reports[29].

†, ++: $p<5\times10^{-8}$, +: $1\times10^{-4}<p<5\times10^{-8}$ with confirmation in other studies, -: no association

Asians, such as *AIRE*, although the reasons for this are unknown.

It is noteworthy that the list of genes includes many T cell receptor (TCR) and costimulatory signal molecules, many NFκB signal molecules and some B-cell-activation molecules, clearly indicating the importance of T and B cells and inflammatory response, especially the NFκB signal pathway. Interestingly, many molecules such as PTPN22, TNFAIP3, CTLA4 and FCRL3 are negative regulators of receptor signaling.

Here we introduce some recently discovered RA-associated genetic polymorphisms.

1)CCR6

CCR6 encodes chemokine receptor 6, which is a surface marker of Th17, a subset of T helper cells producing IL-17. We identified that genetic variation of *CCR6* is associated with RA (p=7.7x10^{-19}, OR=1.19) in Japanese by the combination of GWAS and replication studies[25]. *CCR6* genetic polymorphism is also associated with RA in Caucasians (p=1.5x10^{-11}, OR=1.11)[27]. It is interesting that not only the identified marker SNP (rs3093024) but also the functional dinucleotide polymorphism (rs968334 and the adjacent new SNP: CA, CG and TG variants, TA was not detected) was found to be associated with CCR6 expression (CA< CG<TG) and serum IL-17 level. This is quite an important finding in that Th17 involvement in the RA pathogenesis was supported genetically because there are some arguments that Th17 is not as important in human RA as in the mouse arthritis models[30, 31]. *CCR6* variant is more strongly associated with ACPA (+) RA and is also associated with Graves' disease and Crohn's disease.

2)AIRE

AIRE is a key regulatory molecule of self-antigen presentation in medullary thymic epithelial cells (mTEC). *AIRE* knockout mice lack expression of organ-specific peripheral antigens (e.g. insulin, salivary protein 1, type II collagen) in the mTEC of thymus, which leads to the development of organ-specific autoimmune diseases[32]. Combination of GWAS and replication studies in Japan revealed that genetic polymorphisms of the *AIRE* gene are associated with RA[33]. There were two SNPs with genome-wide significance, one of which is located in an intron and correlated with the decreased expression of *AIRE* gene. This is in concordance with *AIRE* knockout mice developing more rapid and severe collagen-induced arthritis[34]. The other SNP is located in exon 7, which introduces amino acid alteration (S278R) at the SAND domain, and these two SNPs are in strong linkage disequilibrium. Such altered AIRE molecule may have reduced AIRE function.

3)MBP

MBP encodes myelin basic protein, which is a constituent of the myelin sheath of peripheral nerves. We conducted GWAS and replication studies with 2 different cohorts and identified *MBP* as a susceptibility gene for RA[26]. We also found that ~70% of RA patients have anti-MBP antibody in the serum. This was surprising because MBP is an autoantigen for multiple sclerosis (MS) and RA patients do not show such neurological symptoms as MS patients do. However, soon we found that this is not so surprising. First, MBP has several isoforms and the long isoform of MBP is called Golli-MBP[35, 36]. Identified SNP is located in the intron of *Golli-MBP*. *Golli-MBP* is expressed in the hematopoietic cells and was shown to function as a negative regulator of TCR signaling through PKC[37]. *Golli-MBP* knockout T cells showed stronger reaction than the wild-type T cells[38]. Moreover, we found that anti-MBP antibody in the sera of RA recognized citrullinated MBP, but not non-citrullinated MBP. Since MBP is a well-known antigen that is physiologically citrullinated and a number of citrullinated proteins are the targets of RA autoantibodies[39], it is not surprising that MBP becomes one of the targets of RA autoimmunity. However, it has not been well studied how the *MBP* polymorphism is linked to the pathogenesis of RA. The *MBP* polymorphism is not associated with RA in Caucasians.

4)TNFAIP3

The *TNFAIP3* gene encodes a cytoplasmic zinc finger protein that possesses both ubiquitination and deubiquitination properties and is a major negative regulator of TNF-induced NF-κB signaling pathways. *TNFAIP3* polymorphism showed relatively high odds ratio for RA in both Caucasians and Japanese (odds ratios of 1.22 and 1.35, respectively). Several different polymorphisms have been associated with autoimmunity, including a nonsynonymous coding SNP (Phe127Cys), with some evidence of reduced negative regulatory ability for TNF-induced NF-κB signaling[40]. In addition to *TNFAIP3*, a number of genes related to NF-κB signaling (e.g. *TRAF1*, *CD40*, *Rel* and

NFKBIE) were reported to be associated with RA, clearly indicating the importance of NF-κB signaling in the pathogenesis of RA.

In the near future: rare variants

The genetic influence of each polymorphism is very modest (OR mostly ranging from 1.1 to 1.5). Therefore, there is no obvious clinical utility to predict the development of RA with such polymorphisms. This may change as the obtained knowledge becomes more complete, but currently all the known genetic variants can explain only ~15% of the genetic component[41]. This will not change very much even though we have found >100 associated genes with common variants (SNPs). Since most of the GWASs adopt common SNPs with a population prevalence of >3-5%, there may be some rare genetic variants with high genetic impacts. Sialic acid acetylesterase (*SIAE*) is an enzyme that negatively regulates B lymphocyte antigen receptor signaling and is required for the maintenance of immunological tolerance. By sequencing the *SIAE* exons, various defective variants were found in various autoimmune diseases including RA[42]. Defective variants were found in only 2 out of 648 (0.3%) healthy European subjects, whereas 24 out of 923 (2.6%) autoimmune disease patients had defective variants (OR=8.62). The odds ratio for RA was 8.31. Although this result was not successfully replicated in a larger study[43], some unknown rare variants may have strong impacts on the development of RA.

Now that the sequencing technology has developed markedly and is becoming less expensive, finding rare genetic variants associated with RA by whole-genome sequencing is realistic. As a first step, researchers started sequencing only exons of the whole genome, which is called the exome sequence, because it is much more economical than whole-genome sequencing. However, in the very near future, it is announced that the whole-genome sequence of one person can be read for $1,000 in a day. From this point onwards, it will be more realistic to understand completely the impact of genetic variants on the development of RA.

Acknowledgements

We would like to appreciate Prof. Matsuda of Dept. of Genomic Medicine in Kyoto University for all the supports on the research of our genetic studies. We would also like to appreciate all the members of GARNET consortium for the meta-analysis of GWAS and the HLA data on the Japanese RA. This work was supported by Grants-in-aid from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Sciences and Technology of Japan. We have no conflict of interests to be declared.

References

1) Bax M, van Heemst J, Huizinga TW, Toes RE: Genetics of rheumatoid arthritis: what have we learned? Immunogenetics. 2011; 63: 459-466.
2) Gregersen PK, Olsson LM: Recent advances in the genetics of autoimmune disease. Annu Rev Immunol. 2009; 27: 363-391.
3) van Venrooij WJ, van Beers JJ, Pruijn GJ: Anti-CCP antibodies: the past, the present and the future. Nat Rev Rheumatol. 2011; 7: 391-398.
4) Kroot EJ, de Jong BA, van Leeuwen MA, et al: The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. Arthritis Rheum. 2000; 43: 1831-1835.
5) Kallberg H, Padyukov L, Plenge RM, et al: Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet. 2007; 80: 867-875.
6) Padyukov L, Seielstad M, Ong RT, et al: A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. Ann Rheum Dis. 2011; 70: 259-265.
7) Gregersen PK, Silver J, Winchester RJ: The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987; 30: 1205-1213.
8) Deighton CM, Walker DJ, Griffiths ID, Roberts DF: The contribution of HLA to rheumatoid arthritis. Clin Genet. 1989; 36: 178-182.
9) de Vries N, Tijssen H, van Riel PL, van de Putte LB: Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67-74 of the HLA-DRB1 molecule. Arthritis Rheum. 2002; 46: 921-928.
10) Freed BM, Schuyler RP, Aubrey MT: Association of the HLA-DRB1 epitope LA(67, 74) with rheumatoid arthritis and citrullinated vimentin binding. Arthritis Rheum. 2011; 63: 3733-3739.
11) Mattey DL, Dawes PT, Gonzalez-Gay MA, et al: HLA-DRB1 alleles encoding an aspartic acid at position 70 protect against development of rheumatoid arthritis. J Rheumatol. 2001; 28: 232-239.

12) Raychaudhuri S, Sandor C, Stahl EA, et al: Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012; 44: 291-296.
13) Ohmura K, Terao C, Maruya E, et al: Anti-citrullinated peptide antibody-negative RA is a genetically distinct subset: a definitive study using only bone-erosive ACPA-negative rheumatoid arthritis. Rheumatology (Oxford). 2010; 49: 2298-2304.
14) Ding B, Padyukov L, Lundstrom E, et al: Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. Arthritis Rheum. 2009; 60: 30-38.
15) Verpoort KN, van Gaalen FA, van der Helm-van Mil AH, et al: Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. Arthritis Rheum. 2005; 52: 3058-3062.
16) Vignal C, Bansal AT, Balding DJ, et al: Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci. Arthritis Rheum. 2009; 60: 53-62.
17) Lundstrom E, Kallberg H, Smolnikova M, et al: Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum. 2009; 60: 924-930.
18) Terao C, Ohmura K, Kochi Y, et al: A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects. Ann Rheum Dis. 2011; 70: 2134-2139.
19) Mackie SL, Taylor JC, Martin SG, et al: A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. Genes Immun. 2012; 13: 120-128.
20) Suzuki A, Yamada R, Chang X, et al: Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet. 2003; 34: 395-402.
21) Consortium TWTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447: 661-678.
22) Plenge RM, Seielstad M, Padyukov L, et al: TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med. 2007; 357: 1199-1209.
23) Plenge RM, Cotsapas C, Davies L, et al: Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet. 2007; 39: 1477-1482.
24) Gregersen PK, Amos CI, Lee AT, et al: REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet. 2009; 41: 820-823.
25) Kochi Y, Okada Y, Suzuki A, et al: A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet. 2010; 42: 515-519.
26) Terao C, Ohmura K, Katayama M, et al: Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis--a genome-wide study combined with immunological analyses. PLoS One. 2011; 6: e20457.
27) Stahl EA, Raychaudhuri S, Remmers EF, et al: Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010; 42: 508-514.
28) Zhernakova A, Stahl EA, Trynka G, et al: Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011; 7: e1002004.
29) Okada Y, Terao C, Ikari K, et al: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet. 2012; 44: 511-516.
30) Kato H, Fox DA: Are Th17 cells an appropriate new target in the treatment of rheumatoid arthritis? Clin Transl Sci. 2010; 3: 319-326.
31) Genovese MC, Durez P, Richards HB, et al: One year efficacy and safety results of a phase II trial of secukinumab in patients with rheumatoid arthritis. Arthritis Rheum. 2011; 63: S149-S150.
32) Anderson MS, Venanzi ES, Klein L, et al: Projection of an immunological self shadow within the thymus by the aire protein. Science. 2002; 298: 1395-1401.
33) Terao C, Yamada R, Ohmura K, et al: The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Hum Mol Genet. 2011; 20: 2680-2685.
34) Campbell IK, Kinkel SA, Drake SF, et al: Autoimmune regulator controls T cell help for pathogenetic autoantibody production in collagen-induced arthritis. Arthritis Rheum. 2009; 60: 1683-1693.
35) Pribyl TM, Campagnoni CW, Kampf K, et al: The hu-

man myelin basic protein gene is included within a 179-kilobase transcription unit: expression in the immune and central nervous systems. Proc Natl Acad Sci USA. 1993; 90: 10695-10699.

36) Feng JM: Minireview: expression and function of golli protein in immune system. Neurochem Res. 2007; 32: 273-278.

37) Feng JM, Fernandes AO, Campagnoni CW, Hu YH, Campagnoni AT: The golli-myelin basic protein negatively regulates signal transduction in T lymphocytes. J Neuroimmunol. 2004; 152: 57-66.

38) Feng JM, Hu YK, Xie LH, et al: Golli protein negatively regulates store depletion-induced calcium influx in T cells. Immunity. 2006; 24: 717-727.

39) Conrad K, Roggenbuck D, Reinhold D, Dorner T. Profiling of rheumatoid arthritis associated autoantibodies. Autoimmun Rev. 2010; 9: 431-435.

40) Musone SL, Taylor KE, Lu TT, et al: Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet. 2008; 40: 1062-1064.

41) Raychaudhuri S: Recent advances in the genetics of rheumatoid arthritis. Curr Opin Rheumatol. 2010; 22: 109-118.

42) Surolia I, Pirnie SP, Chellappa V, et al: Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. Nature. 2010; 466: 243-247.

43) Hunt KA, Smyth DJ, Balschun T, et al: Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. Nat Genet. 2012; 44: 3-5.

44) Raychaudhuri S, Thomson BP, Remmers EF, et al: Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet. 2009; 41: 1313-1318.

45) Raychaudhuri S, Remmers EF, Lee AT, et al: Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet. 2008; 40: 1216-1223.

46) Daha NA, Kurreeman FA, Marques RB, et al: Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Arthritis Rheum. 2009; 60: 1255-1260.

47) Barton A, Thomson W, Ke X, et al: Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet. 2008; 40: 1156-1159.

48) Kochi Y, Yamada R, Suzuki A, et al: A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet. 2005; 37: 478-485.

49) Suzuki A, Yamada R, Kochi Y, et al: Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. Nat Genet. 2008; 40: 1224-1229.