1A). We vigorously searched candidate markers in exons of *PLD4* which showed strong linkage disequilibrium with rs2841277 and picked up two markers registered in 1000 Genomes Project(34) which displayed more than 5% frequency in genotyped subjects, namely, rs2841280 and rs894037 in exon2 (Figure 1B). Genotyping of these polymorphisms revealed strong linkage disequilibrium between rs2841280 (E27Q) and rs2841277 (D':0.98, r^2:0.75) and monomorphism of rs894037 in a Japanese population. Association study of rs2841280 using control genotypes obtained by imputation supported association of *PLD4* with SSc ($p=6.3x10^{-5}$, Supplementary Table 1 and 2, see methods for detail).

As the three loci are associated with RA in a Japanese population, we analyzed whether the associations with SSc in the current study were brought by patients who suffered from both RA and SSc. When 22 patients who had RA as well as SSc were excluded, we still observed significant associations for the three loci in our study (Table 2). A further stringent analysis excluding subjects with other autoimmune diseases demonstrated significant associations of the three genes (Supplementary Table 2). When we compared SSc with and without other autoimmune diseases for the associated alleles, we did not find any difference (data not shown).

Subanalysis of types of systemic sclerosis

Previous studies have revealed that the genetic background of systemic sclerosis varies between different types of SSc(11, 18). Thus, we performed subanalyses of the five regions examined in the combined study in which we compared the allele frequencies of the controls with those of the lcSSc or dcSSc patients. We used the same controls as in the 1st study or combined study. While *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).

It is also interesting to search whether the susceptibility loci affect autoantibody status and severe complications. The association studies revealed *TNFAIP3* association with SSc patients who possess anti-centromere antibody (ACA) (Supplementary Table 3), but intra-cases study did not demonstrate clear significance (p=0.043). We did not find other associations between the susceptibility loci and clinical phenotypes of SSc both in case-control analyses and intra-case analyses.

Efficacy of the current study

In the current study, we performed candidate gene analysis based on a

meta-analysis of RA GWAS as many autoimmune disease susceptibility genes have been reported to be shared by a wide range of diseases. As a result, we found three susceptibility genes for Japanese systemic sclerosis. Thus, we analyzed whether the candidate gene approach taken in the current study for detecting novel susceptibility genes for SSc was effective. When we calculated the likelihood of finding three susceptibility genes among eighteen genes by chance, we found that it was 2.5×10^{-8} (see method). Our results indicate that our approach to finding 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. While the three susceptibility loci to SSc shared risk direction with RA, we did not detect correspondence of the risk directions of the markers between the two diseases (Figure 2). This indicates that a large proportion of the eighteen RA markers are not shared by SSc and that the lack of association between the thirteen of the markers and SSc was not due to the low power produced by the relatively small number of SSc patients included in this study.

DISCUSSION

As SSc can lead to severe complications, a low quality of life, and short 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 fifteen 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. As a relatively large proportion of RA susceptibility genes is shared by other autoimmune diseases (24), a candidate gene approach using novel markers reported in GWAS of RA is a fascinating way of identifying new SSc markers. In fact, some of the novel susceptibility markers for RA reported in the meta-analysis were shown to be susceptibility markers for systemic lupus erythematosus and Grave's disease (31). In the current study, we successfully found three susceptibility genes for Japanese SSc. No reports have identified *PLD4* as an SSc-associated locus. The current study is also the first report to detect *TNFAIP3* and *IRF8* as susceptibility genes for Japanese SSc. We provided the best fitting model for each association in Supplementary Table 4.

One may argue that these three associations might have been obtained due to the overlap of RA upon SSc. Even after excluding the patients with RA and SSc overlap

based on physicians' reports, we still see the significant associations for the three loci (Table 3). RF and ACPA information was available for 371 and 65 SSc patients without RA, respectively, of which 21.6% and 10.8% were positive for RF and ACPA, respectively. These prevalence are compatible with those reported 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 three loci. These results indicate that the associations in the three loci are not due 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 towards 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 two loci with SSc in a Japanese population.

As *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 linkage disequilibrium (LD) with rs5029939 (r^2=0.85) in the Japanese population. Rs6932056 also displays strong LD with rs2230926, a missense mutation of *TNFAIP3* (r^2=0.85), in the Japanese population. The rs2230926 missense mutation leads to an amino acid alteration in the OUT domain of the A20 protein, which is considered to result in decreased NFk-B signaling. As we did not find strong associations between rs6932056 and SSc in the replication study, it will be necessary to reexamine the association between *TNFAIP3* and SSc using independent Japanese SSc sample sets, 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. While the functions of *PLD4* are also poorly understood, it is known to be involved in the

PLD4 association with Japanese systemic sclerosis

phagocytosis of microglia (38). The expression of *PLD4* around the marginal zone in the spleen might support the functional involvement of *PLD4* in immunological systems. It is pretty interesting that rs2841280, which alters amino acid of PLD4 is associated with SSc. Minor allele G of rs2841280 is associated in a protective manner. The impact of amino acid alteration brought by rs2841280 on effect of PLD4 protein is not known. When we analyzed the impact of the amino acid alteration by SIFT software in-silico(http://sift.jcvi.org/), it was shown to result in a small effect. However, the association raises possibility that this polymorphism leads functional modulation of PLD4 and it is feasible to analyze the functional change of PLD4 protein with rs2841280, using animal models for 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 two genotypes and *PLD4* transcription ($p>0.05$)(39). Taken together, in spite of association of these two mutations, whether one of these two polymorphisms is the causative mutation is not confirmed. While the detection of a p-value of less than $5x10^{-8}$ 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 find 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 two polymorphisms are only markers and that a rare variant in linkage disequilibrium with the two markers affects the disease onset. A causative rare 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). IRF8 protein is a transcription factor involved in interferon pathway. Interferon pathway is shown to be involved with 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. While a European GWAS of SSc patients found suggestive associations between the *IRF4* locus and SSc, it was 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 we classified the SSc patients into lcSSc and dcSSc groups and subjected them to subanalyses, *ARID5B*, *IRF8*, and *CD83* displayed stronger associations with one of the two phenotypes. However, the associations of these three markers with the phenotypes were not strong enough to provide convincing evidence of a clear distinction between the genetic backgrounds of the two SSc phenotypes. When we analyzed the associations of the SSc subtypes with the other thirteen markers in the first set, we did not detect any strong associations ($p>0.05$). Other subanalyses of the susceptibility loci in the combined set did not show significant results between disease phenotypes due to lack of power. As classification according to disease phenotypes resulted in limited number of subjects in each subset, we conducted this subanalyses only in the combined set. The association between *TNFAIP3* and ACA should be confirmed in a large-scale association study.

Although GWAS are an extremely powerful way of detecting novel susceptibility genes for diseases, GWAS of SSc patients have only been performed 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 similar pathological signaling pathways or mechanisms to the disease being examined is an effective way of finding novel susceptibility genes.

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

ACKNOWLEDGEMENTS

This study was performed with the support of the Genetics and Allied research in Rheumatic diseases Networking (GARNET) consortium. We thank the staffs of the BioBank Japan Project for the DNA collection of control subjects.

PLD4 association with Japanese systemic sclerosis

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(4):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(2):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(1):2-10.
4. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. Nat Rev Rheumatol. 2012;8(1):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(2 Suppl 65):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(6):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(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(5):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(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(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(11):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

PLD4 association with Japanese systemic sclerosis

pulmonary fibrosis. Arthritis Rheum. 2009;60(1):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(12):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(4):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(4):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(4):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(2):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(11):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(8):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(6):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(3):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(7):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

PLD4 association with Japanese systemic sclerosis

replication in North Americans. Arthritis Rheum. 2009;60(12):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(23):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(12):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(10):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(12):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(12):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(2):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(3):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(5):511-6.

32. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum. 1980;23(5):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(2):202-5.

34. Consortium GP. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):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(4):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(1):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(11):e13932.

38. Otani Y, Yamaguchi Y, Sato Y, Furuichi T, Ikenaka K, Kitani H, et al. PLD$ is involved in phagocytosis of microglia: expression and localization changes of PLD4 are correlated with activation state of microglia. PLoS One. 2011;6(11):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(5813):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(10):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(11):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(21):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(2):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(1):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(2):127-38.

PLD4 association with Japanese systemic sclerosis

FIGURE LEGENDS

Figure 1. LD block around *PLD4* region and *PLD4* structure.

A.LD block and genes around *PLD4* are shown in the figure. LD block is drawn based on hapmap phase 3 data. Star indicates rs2841277.

B.Schematic view of *PLD4* structure. Rectangles represent exons of *PLD4*.

Figure 2. Comparison of associations between SSc and RA.

The odds ratios obtained for eighteen genes in association studies of SSc and RA are plotted.

PLD4 association with Japanese systemic sclerosis

Table 1. Study population.

		Case	Cont
1st set	Number	415	16,891
	Institutions	Kyoto, TWMU	Kyoto, TWMU, BBJP
	Typing	Taqman Assay	Illumina Human Hap610-Quad BeadChip, Illumina Human Hap550 BeadChip, Affymetrix Genome-wide Human SNP Array 6.0
	Limited%/Diffuse%	49.6%/50.4%	N.A.
	topoI/ACA	30.6%/31.1%	N.A.
	ILD	48.9%	N.A.
	Age	50.9 ± 14.7	60.9 ± 12.5
	Female ratio	91.3%	44.9%
Replication set	Number	315	21,054
	Institutions	Keio, Sagamihara, Kanazawa	Kyoto, BBJP
	Typing	Taqman Assay	Illumina Human Hap550 BeadChip, Illumina Human Hap610-Quad BeadChip
	Limited%/Diffuse%	63.8%/34.6%	N.A.
	topoI/ACA	29.5%/35.2%	N.A.
	ILD	43.2%	N.A.
	Age	51.4 ± 14.1	59.3 ± 14.2
	Female ratio	87.3%	48.4%

N.A.: not applicable, BBJP: BioBank Japan, TWMU: Tokyo Women's Medical University, Kyoto: Kyoto University, Keio: Keio University, Sagamihara: Sagamihara National Hospital, Kanazawa: Kanazawa University, topoI: anti-topoisomerase I antibody, ACA: anti-centromere antibody, ILD: interstitial lung disease

PLD4 association with Japanese systemic sclerosis

Table 2. Association studies of Japanese SSc patients

rsID	Chr	Gene	Allele 1/2	1st set			Replication			Combined study		
				Allele 1 frequency		p-value	Allele 1 frequency		p-value	p-value	OR (95%CI)	p_{nonRA}**
				Control	Case		Control*	Case				
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						

*The control rs6932056 genotypes used in the replication study were imputed using genome-scanning data from 3,765 subjects.

** p-values obtained from comparisons between SSc patients without overlapping RA and controls

PLD4 association with Japanese systemic sclerosis

Table 3. Associations of the two SSc subtypes

rsID	Chr	Gene	Allele 1/2	Control Allele 1freq	lcSSc (N=408)			dcSSc (N=318)		
					Allele 1 freq	p-value	OR (95%CI)	Allele 1 freq	p-value	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)

PLD4 association with Japanese systemic sclerosis

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)

PLD4 association with Japanese systemic sclerosis

Figures

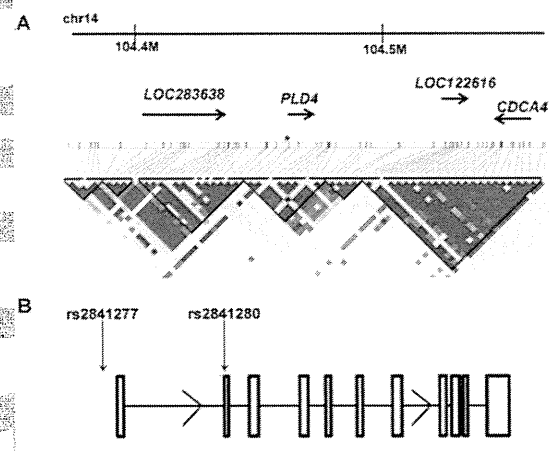

Figure 1. LD block around *PLD4* region and *PLD4* structure.

A.LD block and genes around *PLD4* are shown in the figure. LD block is drawn based on hapmap phase 3 data. Star indicates rs2841277.

B.Schematic view of *PLD4* structure. Rectangles represent exons of *PLD4*.

PLD4 association with Japanese systemic sclerosis

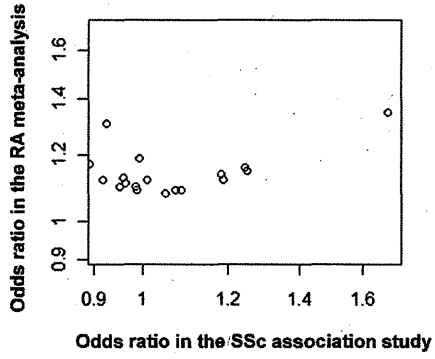

Figure 2. Comparison of associations between SSc and RA. The odds ratios obtained for eighteen genes in association studies of SSc and RA are plotted.

OPEN ACCESS Freely available online PLoS GENETICS

A Genome-Wide Association Study Identified *AFF1* as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese

Yukinori Okada[1,2,3☯], Kenichi Shimane[1,2☯], Yuta Kochi[1,2☯*], Tomoko Tahira[4], Akari Suzuki[1], Koichiro Higasa[3], Atsushi Takahashi[3], Tetsuya Horita[5], Tatsuya Atsumi[5], Tomonori Ishii[6], Akiko Okamoto[2], Keishi Fujio[2], Michito Hirakata[7], Hirofumi Amano[8], Yuya Kondo[9], Satoshi Ito[9], Kazuki Takada[10], Akio Mimori[11], Kazuyoshi Saito[12], Makoto Kamachi[13], Yasushi Kawaguchi[14], Katsunori Ikari[14], Osman Wael Mohammed[15], Koichi Matsuda[15], Chikashi Terao[16,17], Koichiro Ohmura[16], Keiko Myouzen[1], Naoya Hosono[18], Tatsuhiko Tsunoda[19], Norihiro Nishimoto[20], Tsuneyo Mimori[16], Fumihiko Matsuda[17], Yoshiya Tanaka[12], Takayuki Sumida[9], Hisashi Yamanaka[14], Yoshinari Takasaki[8], Takao Koike[5], Takahiko Horiuchi[21], Kenshi Hayashi[4], Michiaki Kubo[18], Naoyuki Kamatani[3], Ryo Yamada[1,17], Yusuke Nakamura[15], Kazuhiko Yamamoto[1,2]

1 Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan, **2** Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, **3** Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan, **4** Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan, **5** Department of Medicine II, Hokkaido University Graduate School of Medicine, Sapporo, Japan, **6** Department of Hematology and Rheumatology, Tohoku University Graduate School of Medicine, Sendai, Japan, **7** Division of Rheumatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan, **8** Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine, Tokyo, Japan, **9** Division of Clinical Immunology, Doctoral Program in Clinical Sciences, Graduate School of Comprehensive Human Science, University of Tsukuba, Tsukuba, Japan, **10** Departments of Medicine and Rheumatology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan, **11** Division of Rheumatic Diseases, National Center for Global Health and Medicine, Tokyo, Japan, **12** First Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan, **13** Department of Immunology and Rheumatology, Unit of Translational Medicine, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan, **14** Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, **15** Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, **16** Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, **17** Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **18** Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan, **19** Laboratory for Medical Informatics, CGM, RIKEN, Yokohama, Japan, **20** Laboratory of Immune Regulation, Wakayama Medical University, Wakayama, Japan, **21** Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan

Abstract

Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Although recent genome-wide association studies (GWAS) have contributed to discovery of SLE susceptibility genes, few studies has been performed in Asian populations. Here, we report a GWAS for SLE examining 891 SLE cases and 3,384 controls and multi-stage replication studies examining 1,387 SLE cases and 28,564 controls in Japanese subjects. Considering that expression quantitative trait loci (eQTLs) have been implicated in genetic risks for autoimmune diseases, we integrated an eQTL study into the results of the GWAS. We observed enrichments of cis-eQTL positive loci among the known SLE susceptibility loci (30.8%) compared to the genome-wide SNPs (6.9%). In addition, we identified a novel association of a variant in the AF4/FMR2 family, member 1 (*AFF1*) gene at 4q21 with SLE susceptibility (rs340630; $P=8.3\times10^{-9}$, odds ratio = 1.21). The risk A allele of rs340630 demonstrated a cis-eQTL effect on the *AFF1* transcript with enhanced expression levels ($P<0.05$). As *AFF1* transcripts were prominently expressed in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes, up-regulation of *AFF1* may cause the abnormality in these lymphocytes, leading to disease onset.

Citation: Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, et al. (2012) A Genome-Wide Association Study Identified *AFF1* as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese. PLoS Genet 8(1): e1002455. doi:10.1371/journal.pgen.1002455

Editor: Mark I. McCarthy, University of Oxford, United Kingdom

Received May 16, 2011; **Accepted** November 18, 2011; **Published** January 26, 2012

Funding: This study was supported by a grant from the CGM, RIKEN, and a grant from the autoimmune disease study group of Research in Intractable Diseases, Japanese Ministry of Health, Labor, and Welfare, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ykochi@src.riken.jp

☯ These authors contributed equally to this work.

Author Summary

Although recent genome-wide association study (GWAS) approaches have successfully contributed to disease gene discovery, many susceptibility loci are known to be still uncaptured due to strict significance threshold for multiple hypothesis testing. Therefore, prioritization of GWAS results by incorporating additional information is recommended. Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Considering that abnormalities in B cell activity play essential roles in SLE, prioritization based on an expression quantitative trait loci (eQTLs) study for B cells would be a promising approach. In this study, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS and identified *AFF1* as a novel SLE susceptibility loci. We also confirmed cis-regulatory effect of the locus on the *AFF1* transcript. Our study would be one of the initial successes for detecting novel genetic locus using the eQTL study, and it should contribute to our understanding of the genetic loci being uncaptured by standard GWAS approaches.

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by autoantibody production, complement activation, and multi-organ damage [1]. Familial aggregation demonstrates that both genetic and environmental factors play a role in pathogenesis of SLE [2]. Genetic studies using candidate gene-approaches, and recently, genome-wide association studies (GWAS), have uncovered more than 25 SLE susceptibility genes, including *HLA-DRB1*, *IRF5*, *STAT4*, *ITGAM*, *BLK*, *TNFAIP3*, and others [3–18]. However, most of these studies were conducted in European populations [3–13,15,17], and few studies have been conducted in Asian populations [14,16,18]. Since the epidemiology of SLE has demonstrated that the prevalence of disease substantially differs among populations, genetic backgrounds of SLE should be also heterogeneous across populations [19,20]. Therefore, additional studies in Asians might provide novel insights. It is of note that GWAS for SLE in Chinese populations identified novel loci that had not been detected in Europeans, such as *ETS1*, *IKZF1*, and *WDFY4* [14,16].

Another issue raised by the previous GWASs for complex diseases is that many susceptibility loci still remained uncaptured, owing to its strict significance threshold for multiple hypothesis testing [21]. In SLE, for example, the 26 risk loci identified by the previous GWAS explained only an estimated 8% of the total genetic susceptibility to the disease [15]. Therefore, it is still important to examine the sub-loci of GWAS, in order to reveal the entire picture of genetic etiology. To effectively explore these uncaptured loci, prioritization of GWAS results by incorporating additional information implicated in the disease pathophysiology is recommended [22,23]. Considering that abnormalities in B cell activity play essential roles in SLE [1] and that expression quantitative trait loci (eQTL) have been implicated to comprise approximately a half of genetic risks for autoimmune diseases [24], prioritization based on an eQTL study for B cells would be a promising approach for SLE [25]. Moreover, an eQTL itself assures the presence of functional variant(s) that regulate gene expression. Thus, eQTL increases the prior probability of the presence of disease-causal variant(s) in the locus more effectively and unbiasedly, compared to other knowledge-based prioritizations such as gene pathway analysis [24].

Here, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS, which effectively enabled to detect a novel SLE susceptibility locus.

Results

GWAS for SLE

In the GWAS, 891 SLE cases and 3,384 controls in Japanese subjects were genotyped over 550,000 single nucleotide polymorphism (SNP) markers (Table S1, S2 and Figure 1). We applied stringent quality control (QC) criteria and evaluated associations of 430,797 autosomal SNPs, as previously described [26]. No substantial population stratification was demonstrated through principal component analysis (Figure S1) or a Quantile–Quantile plot of *P*-values (inflation factor, $\lambda_{GC} = 1.088$, Figure S2), suggesting homogenous ancestries of our study population [27].

We identified significant associations in six chromosomal loci that satisfied the genome-wide significance threshold of $P<5.0\times10^{-8}$ (Table 1 and Figure 2A). These loci have been reported to be associated with SLE susceptibility (*STAT4*, *TNFAIP3*, *HIP1*, *BLK*, *ETS1*, and the HLA region) [3–18]. We also observed significant replications in 17 of the previously reported SLE susceptibility loci [3–18] ($\alpha = 0.01$; Table 2). Of these, significant replications were enriched in the loci identified through the studies in Asian populations (80%; 8 of the 10 loci), including *RASGRP3*, *IKZF1*, *HIP1*, *WDFY4*, intergenic region at 11q23, *ETS1*, *SLC15A4*, *ELF1*, and *HIC2-UBE2L3* [14,16,18], compared to those in European populations (56.3%; 9 of the 16 loci) [3–13,15,17].

Incorporation of eQTL study into GWAS results

For the selection of SNPs incorporated in the replication studies of the potential association signals, we evaluated cis-eQTL effects of the SNPs using publically available gene expression data [28], and prioritized the results of the GWAS. After applying QC criteria, we evaluated the expression levels of 19,047 probes assayed in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals [29] using Illumina's human whole-genome expression array (WG-6 version 1) [28]. For each of the SNPs included in our GWAS, probes located within ±300 kbp regions were focused on as cis-eQTLs (average 4.93 probes per SNP). We denoted the SNPs which exhibited significant associations with expression levels of any of the corresponding cis-eQTLs as eQTL positive (false discovery rate (FDR) *Q*-values<0.2). We observed enrichments of eQTL positive loci among the SLE susceptibility loci (30.8%; 8 of the 26 evaluated loci) including a well-known eQTL gene of *BLK* [11,25] (Table 2), compared to the genome-wide SNPs (6.9%) and compared even to the SNPs in the vicinity of expressed loci (among the SNPs located within ±10 kbp of probes used for the expression analysis, 13.1% were eQTL positive; Table S3).

By prioritizing the results of the GWAS using the eQTL study, we selected 57 SNPs from 1,207 SNPs that satisfied $P<1.0\times10^{-3}$ in the GWAS. We subsequently referred the associations of the selected SNPs using the results of the concurrent genome-wide scan for SLE in an independent Japanese population (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009). In the scan, 447 SLE cases and 680 controls of Japanese origin were evaluated using a pooled DNA approach [30]. We selected SNPs if any association signals were observed in the neighboring SNPs of the

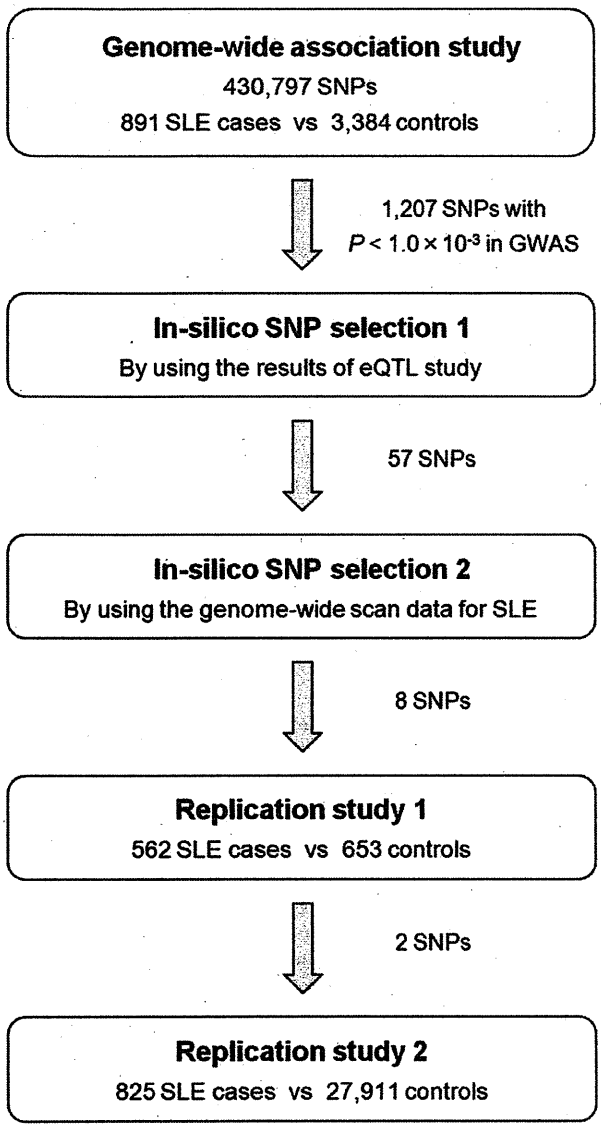

Figure 1. Design of the GWAS and multi-stage replication studies for SLE in Japanese subjects. A total of 2,278 SLE cases and 31,948 controls were enrolled. The clinical characteristics of the subjects are summarized in Table S1 and S2. Details of the genome-wide scan data for SLE referenced in the *in silico* SNP selection 2 are described elsewhere (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009).
doi:10.1371/journal.pgen.1002455.g001

pooled analysis. As a result, 8 SNPs remained for further investigation (Table S4).

Replication studies and identification of *AFF1*

Then, we performed two-stage replication studies using independent SLE cohorts for Japanese subjects (cohort 1 with 562 SLE cases and 653 controls, and cohort 2 with 825 SLE cases and 27,911 controls). First, we evaluated the selected 8 SNPs in the replication study 1. In the replication study 2, 2 SNPs that satisfied $P<1.0\times10^{-6}$ in the combined study of GWAS and replication study 1 were further evaluated (Figure 1). Among the evaluated SNPs, we observed significant replications in the SNP located in the genomic region of the AF4/FMR2 family, member 1 gene (*AFF1*) at 4q21 (rs340630; $P=4.6\times10^{-5}$ and $P=0.0094$ in the two individual cohorts, respectively; Table 3, Table S5, and Figure 2B). The combined study for the GWAS ($P=1.5\times10^{-4}$) and the replication studies demonstrated significant associations of rs340630 that satisfied the genome-wide significance threshold ($P=8.3\times10^{-9}$, OR = 1.21, 95% CI 1.14–2.30).

Cis-eQTL effect of rs340630 on *AFF1* transcripts

Since the landmark SNP in the *AFF1* locus, rs340630, was prioritized through the eQTL study as an eQTL positive SNP (Table 3), we further validated its cis-eQTL effect using Epstein-Barr virus (EBV)-transfected B cell lines established from Japanese individuals (Pharma SNP Consortium (PSC) cells, $n=62$). The correlation between rs340630 genotypes and the expression levels of *AFF1* was significant in the PSC cells stimulated with phorbol myristate acetate (PMA) ($R^2=0.074$, $P=0.033$; Figure 3A). The expression levels increased with the number of SLE-risk (A) alleles. To further confirm this cis-regulatory effect, we performed allele-specific transcript quantification (ASTQ) of *AFF1*. The transcript levels of each allele were quantified by qPCR using an allele specific probe for a SNP in the 5′-untranslated region (rs340638), which was in absolute LD with rs340630 ($r^2=1.0$, $D'=1.0$). We examined PSC-cells ($n=17$) that were heterozygous for both rs340630 and rs340638. The mean ratio of each transcript (A over G allele; the A allele comprises a haplotype with the risk (A) allele of rs340630) were significantly increased to 1.07 compared to the ratio of the amount of DNA (1.00, $P=0.012$) (Figure 3B). These results suggest that rs340630, or SNP(s) in LD with it, are a regulatory variant predisposing SLE susceptibility through increased expression levels of *AFF1*.

Expression of *AFF1* in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes

AFF1 is known to be involved in cytogenetic translocations of acute lymphoblastic leukemia (ALL) [31]. Its fusion protein with the mixed-lineage leukemia gene (*MLL*) is implicated in the regulation of transcription and the cell cycle of lymphocytes [31]. To investigate the expression pattern of *AFF1* in normal tissues, we evaluated the transcript levels of *AFF1* in a panel of various tissues. We observed prominent expression of *AFF1* in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes, implying an important role for *AFF1* in helper-T-cells and B-cells (Figure 3C).

Discussion

Through a GWAS and multi-staged replication studies consisting of 2,278 SLE cases and 31,948 controls in Japanese subjects, our study identified that the *AFF1* locus was significantly associated with SLE susceptibility.

As well as the identification of the novel SLE susceptibility locus, we observed significant replications of associations in the previously reported susceptibility loci. The replications were especially enriched in the loci identified through the studies in Asian populations, compared to those in European populations. Considering the ethnical heterogeneities in the epidemiology of SLE [19,20], these observations suggest the similarities in the genetic backgrounds of SLE shared within Asian populations, and also the existence of the both common and divergent genetic backgrounds encompassed between European and Asian populations.

Table 1. Results of a genome-wide association study for Japanese patients with SLE.

rsID[a]	Chr	Position (bp)	Cytoband	Gene	Allele[b]	No. subjects		Allele 1 freq.		OR (95%CI)	P
					1/2	Case	Control	Case	Control		
rs10168266	2	191,644,049	2q32	*STAT4*	T/C	891	3,384	0.37	0.27	1.59 (1.42–1.78)	2.7×10^{-16}
rs9501626	6	32,508,322	6p21	HLA region	A/C	891	3,381	0.20	0.12	1.86 (1.62–2.13)	1.0×10^{-18}
rs2230926	6	138,237,759	6q23	*TNFAIP3*	G/T	891	3,377	0.11	0.069	1.75 (1.47–2.08)	1.9×10^{-10}
rs6964720	7	75,018,280	7q11	*HIP1*	G/A	891	3,384	0.25	0.19	1.43 (1.27–1.63)	1.3×10^{-8}
rs2254546	8	11,381,089	8p23	*BLK*	G/A	891	3,384	0.78	0.72	1.42 (1.61–1.25)	4.1×10^{-8}
rs6590330	11	127,816,269	11q24	*ETS1*	A/G	891	3,368	0.48	0.39	1.44 (1.30–1.60)	1.3×10^{-11}

[a]SNPs that satisfied the threshold of $P<5.0\times10^{-8}$ were indicated.
[b]Based on forward strand of NCBI Build 36.3.
SLE, systemic lupus erythematosus; OR, odds ratio.
doi:10.1371/journal.pgen.1002455.t001

To effectively detect the novel SLE susceptibility locus, we integrated cis-eQTL effects of the SNPs and prioritized the results of the GWAS. In addition to identifying a novel locus for SLE-susceptibility, our study demonstrated approximately 30% of confirmed SLE-susceptibility loci were comprised of cis-eQTLs. We also confirmed cis-regulatory effect of the landmark SNP in the *AFF1* locus, rs340630, on *AFF1* transcripts, which had been prioritized through the eQTL study. These results would suggest that accumulation of quantitative changes in gene expression would accelerate the disease onset of SLE. It would also demonstrate the validity of applying eQTL study in the search of the susceptible genes for SLE or other autoimmune diseases, as previously suggested in the study for celiac disease [24]. To our knowledge, this is one of the initial studies to successfully discover a new locus by prioritizing GWAS results using eQTLs, and should contribute to the approaches assessing genetic loci still being uncaptured by recent large-scaled GWASs due to stringent significance threshold for multiple hypothesis testing [21].

We observed prominent expression levels of *AFF1* in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes, which would imply an important role for *AFF1* in helper-T-cells and B-cells. In fact, *AFF1* is essential for normal lymphocyte development, as demonstrated in mice deficient for *AFF1*; severe reduction were observed in the thymic double positive CD4/CD8 population and the bone marrow pre-B and mature B-cell numbers [32]. The risk A allele of rs340630 demonstrated a cis-eQTL effect on the *AFF1* transcript with enhanced expression levels. As the *AFF1* locus was also demonstrated as an eQTL in primary liver cells [33], the cis-regulatory effect may hold in primary cells as well as lymphoblastoid cells used in the present study. However, because the mechanism of transcriptional regulation is substantially different among cell types [34], cell-type specific analyses including those for primary T-cells and B-cells are needed for understanding the precise role of *AFF1* variant in primary lymphocytes. Although further functional investigation is necessary, our observation suggested that *AFF1* is involved in the etiology of SLE through the regulation of development and activity of lymphocytes. It is of note that *AFF3*, which also belongs to the AF4/FMR2 family, is associated with susceptibility to autoimmune diseases [35].

One of our study's limitations is the selection of SNPs for the replication study using the results of the pooled DNA approach [30], which used a different genotyping platform from that of the present GWAS. Moreover, the association signals based on Silhouette scores in pooled analysis would be less reliable compared to those based on individual genotyping. Since direct comparisons of the association signals of the same single SNPs between the studies would be difficult due to these issues, we adopted the complementary approach that referred the association signals of the multiple SNPs in the pooled analysis for each of the single SNPs in the GWAS, taking account of LD and physical distances between the SNPs. However, there would exist a possibility that the variant(s) truly associated with SLE was left not to be examined in the replication study. It should be noted that only 1 SNP among the 8 selected SNPs yielded the significant association with SLE, although further enrichments of the significant associations might be anticipated. To elucidate effectiveness and limitation of our approach, further assessments of the studies on the remaining loci would be desirable. It should also be noted that the control-case ratio of the subjects were relatively high in the replication study 2 (=33.8), and this disproportionate ratio could have induced potential bias on the results of the association analysis of the SNPs. However, considering the homogeneous ancestries of the Japanese population [27] and that principal component analysis did not demonstrate significant population stratification in the control subjects of the replication study 2 (data not shown), the bias owing to population stratification might not be substantial.

In summary, through a GWAS and multi-staged replication studies in a Japanese population integrating eQTL study, our study identified *AFF1* as a novel susceptibility locus for SLE.

Materials and Methods

Subjects

We enrolled 2,278 systemic lupus erythematosus (SLE) cases and 31,948 controls. SLE cases enrolled in the genome-wide association study (GWAS) ($n=891$) or part of the 2nd replication study ($n=83$) were collected from 12 medical institutes in Japan under the support of the autoimmune disease study group of Research in Intractable Diseases, Japanese Ministry of Health, Labor and Welfare: Hokkaido University Graduate School of Medicine, Tohoku University Graduate School of Medicine, the University of Tokyo, Keio University School of Medicine, Juntendo University School of Medicine, University of Occupational and Environmental Health, University of Tsukuba, Tokyo Medical and Dental University, National Center for Global Health and Medicine, Nagasaki University, Wakayama Medical University, and Jichi Medical University. SLE cases ($n=562$) and controls ($n=653$) enrolled in the 1st replication study were collected from Kyushu University. Some of the SLE cases ($n=742$) and controls ($n=27,911$) enrolled in the 2nd replication study were collected from Kyoto University, Tokyo Women's