

Identification of a significant association of a single nucleotide polymorphism in *TNXB* with systemic lupus erythematosus in a Japanese population

Yoichiro Kamatani · Koichi Matsuda · Tetsuya Ohishi · Shigeru Ohtsubo · Keiko Yamazaki · Aritoshi Iida · Naoya Hosono · Michiaki Kubo · Wako Yumura · Kosaku Nitta · Toyomasa Katagiri · Yasushi Kawaguchi · Naoyuki Kamatani · Yusuke Nakamura

Received: 29 September 2007 / Accepted: 22 October 2007 / Published online: 6 December 2007
© The Japan Society of Human Genetics and Springer 2007

Abstract Systemic lupus erythematosus (SLE) is one of the common autoimmune diseases, with complex genetic components. Here, we report on a case–control association study of 178 SLE patients and 899 control subjects, using genome-wide gene-based single nucleotide polymorphism (SNP) markers. An SNP, rs3130342, in a 5' flanking region of the *TNXB* gene revealed a significant association with

SLE [$P = 0.000000930$, odds ratio (OR) 3.11, with 95% confidence interval (95%CI) of 1.89–5.28] in a Japanese population. This association was replicated independently with 203 cases and 294 controls ($P = 0.0440$, OR 1.52, with 95%CI of 1.01–2.78). Although a copy number variation (CNV) of the *C4* gene adjacent to the *TNXB* gene was reported to be associated with SLE, our analysis on this CNV revealed that the association of CNV of the *C4* gene was weaker than the SNP in the *TNXB* gene and likely to reflect the linkage disequilibrium between *C4* CNV and this particular SNP. Stratified analysis also revealed that the association of SNP rs3130342 with SLE was independent of the HLA-DRB1*1501 allele that has been shown to be associated with SLE. Our findings strongly imply that the *TNXB* gene is a candidate gene susceptible to SLE in the Japanese population.

Electronic supplementary material The online version of this article (doi:10.1007/s10038-007-0219-1) contains supplementary material, which is available to authorized users.

Y. Kamatani · K. Matsuda (✉) · T. Ohishi · S. Ohtsubo · T. Katagiri · Y. Nakamura
Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo, 4-6-1 Shirokanedai, Minato, Tokyo 108-8639, Japan
e-mail: koichima@ims.u-tokyo.ac.jp

T. Ohishi · S. Ohtsubo · W. Yumura · K. Nitta
Department of Medicine, Kidney Center,
Tokyo Women's Medical University, Tokyo, Japan

K. Yamazaki
Laboratory for Gastrointestinal Diseases,
SNP Research Center, The Institute of Physical
and Chemical Research (RIKEN), Kanagawa, Japan

A. Iida
Laboratory for Pharmacogenetics, SNP Research Center,
The Institute of Physical and Chemical Research (RIKEN),
Tokyo, Japan

N. Hosono · M. Kubo
Laboratory for Genotyping, SNP Research Center,
The Institute of Physical and Chemical Research (RIKEN),
Kanagawa, Japan

Y. Kawaguchi · N. Kamatani
Institute of Rheumatology,
Tokyo Women's Medical University, Tokyo, Japan

Keywords Systemic lupus erythematosus · single nucleotide polymorphism (SNP) · *TNXB* · Copy number variation (CNV) · *C4* · Genome-wide association study

Introduction

Systemic lupus erythematosus (SLE) (OMIM #152700) is a complex, heterogeneous, and relatively prevalent autoimmune disease in Japan. Glucocorticoids, together with immunosuppressive agents, have been widely used for treatment of SLE patients and have remarkably improved the clinical outcome for SLE patients (Trager and Ward 2001). However, most of the patients with SLE need to take these drugs throughout their lives after the diagnosis and often suffer from adverse events, such as opportunistic infections, avascular necrosis or myocardial infarction.

Even in patients for whom the disease's activity could be successfully controlled at an early stage, progression of diseases such as renal insufficiency and neurological complications might afflict the patients later in the course. Hence, it is crucial for its detailed pathogenesis to be elucidated and novel therapeutic modalities developed that will improve patients' prognoses.

Although the molecular mechanisms of SLE have not been fully clarified, the involvement of both environmental and genetic factors in the etiology of SLE has been implied. Concordance rate of SLE in monozygote twins (24–58%) was significantly higher than that in dizygote twins (less than 10%), and a sibling recurrence risk ratio (λ_s) of the disease was estimated to be ~ 20 (Croker and Kimberly 2005). Therefore, the presence of multiple genetic factors that contribute to its pathogenesis has been suspected. To investigate genes susceptible to SLE, 11 studies applying genome-wide linkage analysis were conducted, and eight loci were shown to have significant linkage to SLE (Tsao 2003). A combined analysis of these studies indicated three loci at 6p21.1-q15, 16p13-q12.2 and 20p11-q13.13 to be candidate regions for SLE susceptibility (Forabosco et al. 2006).

In the very recent years, an information database of very high-density single nucleotide polymorphisms (SNPs) across the genome has been constructed, and advances in genotyping technology permit high-throughput SNP analysis at low cost. Hence, the SNP-based genome-wide approach is now widely used to investigate the genetic factors involved in diseases with high prevalence and low penetrance. By applying this approach, our group has identified genes susceptible to various diseases, such as myocardial infarction, rheumatoid arthritis, osteoarthritis, and cerebral infarction (Ozaki et al. 2002; Suzuki et al. 2003; Seki et al. 2005; Kubo et al. 2007).

In the study described here, we performed a genome-wide case-control association study to exploit genes susceptible to SLE and found that an SNP in the *TNXB* gene within the major histocompatibility complex (MHC) region of chromosome 6p21 had a strong association with SLE. Our genome-wide association study identified an SLE-susceptible locus on the MHC region in the Japanese population.

Materials and methods

DNA samples and genotyping

A total of 178 individuals with SLE was diagnosed at the Department of Nephrology, Tokyo Women's Medical University (case 1 and case 2, Supplementary Table 1). Additional DNA samples of 203 individuals with SLE

(case 3) were obtained from the Department of Rheumatology. All patients were diagnosed to have SLE on the basis of the revised criteria by the American College of Rheumatology (Hochberg 1997). Two sets of 538 (control 1) and 365 (control 2) control subjects were recruited, independently, from various medical institutes in Japan. We obtained written informed consent from each subject, and this study was approved by the ethics committee at the Institute of Medical Sciences, University of Tokyo, as well as that at Tokyo Women's Medical University. We extracted genomic DNA from peripheral blood leukocytes, using standard protocols. The third control cohort (control 3) was a gender- and age-matched control consisting of 294 healthy volunteers collected by the Pharma SNP Consortium (Ikari et al. 2006). We genotyped SNPs using the multiplex polymerase chain reaction (PCR)-based Invader assay (Third Wave Technologies). The detailed method has been described elsewhere (Ohnishi et al. 2001). We sequenced the *TNXB* gene in 48 healthy Japanese control subjects to search genetic variations further. Typing for the HLA-DRB1 locus was performed by the PCR-microtiter plate hybridization (MPH) method using commercial typing kits (WAK Flow, Wakunaga, Hiroshima, Japan), as described previously (Kawai et al. 1996).

Copy number analysis of C4 gene

The *C4* gene has two isoforms, *C4A* and *C4B*, differing by only four amino acids at positions 1,101, 1,102, 1,105, and 1,106 (Supplementary Fig. 1). Another two isoforms, *C4* long (*C4L*) and *C4* short (*C4S*), were distinguished by the presence (*C4L*) or absence (*C4S*) of the insertion of endogenous retrovirus sequence HERV-K (*C4*) in intron 9 (Dangel et al. 1995). The TaqMan probes to evaluate copy numbers of the *C4A* and *C4B* genes have been previously described (Szilagyi et al. 2006). We designed primers and probe for *C4S* by Primer Express software, version 1.5 (Applied Biosystems). Probes and primer sets used in this assay are listed in Supplementary Table 2. We used 5' VIC-labeled *RNaseP* control reagents (Applied Biosystems) as a reference gene probe. We prepared 10 μ l of reaction mixture containing 20 ng of genomic DNA, 900 nM of primers, 250 nM of probes, 0.5 μ l of *RNaseP* control reagent, and 5 μ l of TaqMan Universal Master Mix (Applied Biosystems). Thermal cycling conditions were as follows: initial 2-min incubation at 50°C and 10-min incubation at 95°C, and following the 40-cycle reaction at 95°C for 15 s and 60°C for 1 min. The PCR reaction was performed in a 384-well reaction plate using ABI 7900HT (Applied Biosystems). A genomic fragment of the *RNaseP* gene, together with the target fragment, was co-amplified within the same tube to apply the copy number of the

RNase P gene for normalization. Experiments were at least triplicated for each sample. The real-time amplification data were analyzed by sequence detection system software (version 2.1; Applied Biosystems). Relative gene copy numbers were determined by the comparative C_T method (User Bulletin #2; Applied Biosystems). Outliers within triplicates were excluded by the result of Smirnov–Grubb’s test. We verified the accuracy of this quantification method by comparing the result of quantitative polymerase chain reaction (qPCR) analysis and that of Southern blotting analysis (Supplementary Fig. 2). The total gene copy number of *C4* was estimated from the sum of *C4A* and *C4B*, as well as that of *C4S* and *C4L*. Thus, we determined the copy numbers of total *C4* and *C4L* from those of *C4A*, *C4B*, and *C4S*.

Statistical analysis

A genome-wide association study of SNPs in SLE patients and control subjects had been already carried out. To increase the statistical power, we adopted the two-step approach (Saito and Kamatani 2002). At the first stage, we performed an association study of 94 SLE patients (case 1) and 538 control subjects (control 1) for the genome-wide approach using 52,608 gene-based SNPs (Tsunoda et al. 2004). Then, we chose SNPs that showed a *P* value of 0.01 or less in either of three genetic models and further performed SNP genotyping in 84 SLE patients (case 2) and 365 independent control subjects (control 2). In each stage, SNPs showing a *P* value of 0.01 or lower in the Hardy–Weinberg equilibrium test for control samples were excluded. Fisher’s exact test was applied for the calculation of statistical significance of two-by-two contingency tables in three genetic models: an allele count model, a dominant effect model, and a recessive effect model. We used Haploview software, version 3.32 (Barrett et al. 2004) to draw a linkage disequilibrium (LD) map around the marker SNP. In the replication study, the independent 203 patients (case 3) and 294 controls (control 3) were genotyped for SNP rs3130342, and the analyses were carried out by the same statistical method. The DNA from a total of 178 patients (cases 1 and 2) and 365 controls (control 2) was also analyzed for variations in gene copy number. To investigate the differences in the number of copies between the case and control groups, we used the Wilcoxon rank sum test. Differences in the distribution of copy number among the three genotype groups were analyzed by the Kruskal–Wallis test. Multivariate logistic regression modeling was used to estimate the independency of each polymorphism for SLE susceptibility, concerning SNP rs3130342 and the number of copies of the *C4A* gene and the *C4B* gene. Fisher’s exact test was also used for the test

of association of HLA-DRB1*1501 with SLE by allele count model. For statistical analysis, we used the R statistical software version 2.4.1.

Results

To screen for a gene(s) susceptible to SLE, we performed genome-wide, case–control, association analysis using 52,608 gene-based SNPs for 94 SLE patients (case 1) and 538 controls (control 1). Among the 50,464 SNPs successfully genotyped, 1,310 showed *P* values of 0.01 or smaller. Subsequently, we genotyped for the additional 84 SLE patients (case 2) and 365 controls (control 2) for these 1,310 SNPs, and found an SNP, rs1009382, to have a strong association with SLE, with a *P* value of 0.00894, in the second stage; the combined samples at allele count model revealed a *P* value of 0.00000518 (Supplementary Table 3). This *P* value is the second significant *P* value obtained by this approach following an SNP in the *ITPR3* gene recently identified (Ohishi et al., data not shown).

Next, we constructed an LD map around this marker SNP on the basis of the HapMap-JPT genotyping data [release 21, based on National Center for Biotechnology Information (NCBI) build 35 and single nucleotide polymorphism database (dbSNP) build 125] using SNPs with the minor allele frequency of 20% or higher. As shown in Fig. 1b, two LD blocks were identified in the genomic region, including the SNP rs1009382. We selected 45 tag-SNPs from the genomic region covering these LD blocks and genotyped them for 178 patients (cases 1 and 2) and 365 controls (control 2). The statistical analysis indicated that the association was still peaked at this marker SNP rs1009382 in exon 23 of the *TNXB* gene and became gradually weak in both directions (Fig. 1a), implying that *TNXB* was the most likely candidate gene susceptible to SLE.

To define further the region of interest, we screened SNPs in a genomic region corresponding to the *TNXB* gene and identified 60 SNPs. Since the CNV region was located adjacent to this locus and contained a part of the *TNXB* gene, we selected and genotyped 33 SNPs mapped to a single-copy part of the human genome and found that the SNP rs3130342 located in the 5’ flanking region of the *TNXB* gene revealed a stronger association [*P* value of 0.000000930, with odds ratio (OR) of 3.11 and 95% confidence interval (95%CI) of 1.89–5.28] than did the marker SNP (Table 1). Although the *ITPR3* gene, for which we had previously revealed a significant association with SLE, is located in the same chromosomal region as the *TNXB* gene is, the distance between the two genes is 1.5 Mb, and the pairwise *D'* value between them is very small. Hence, the association of the SNP on the *TNXB* gene with SLE was considered to be independent of that on the *ITPR3* gene.

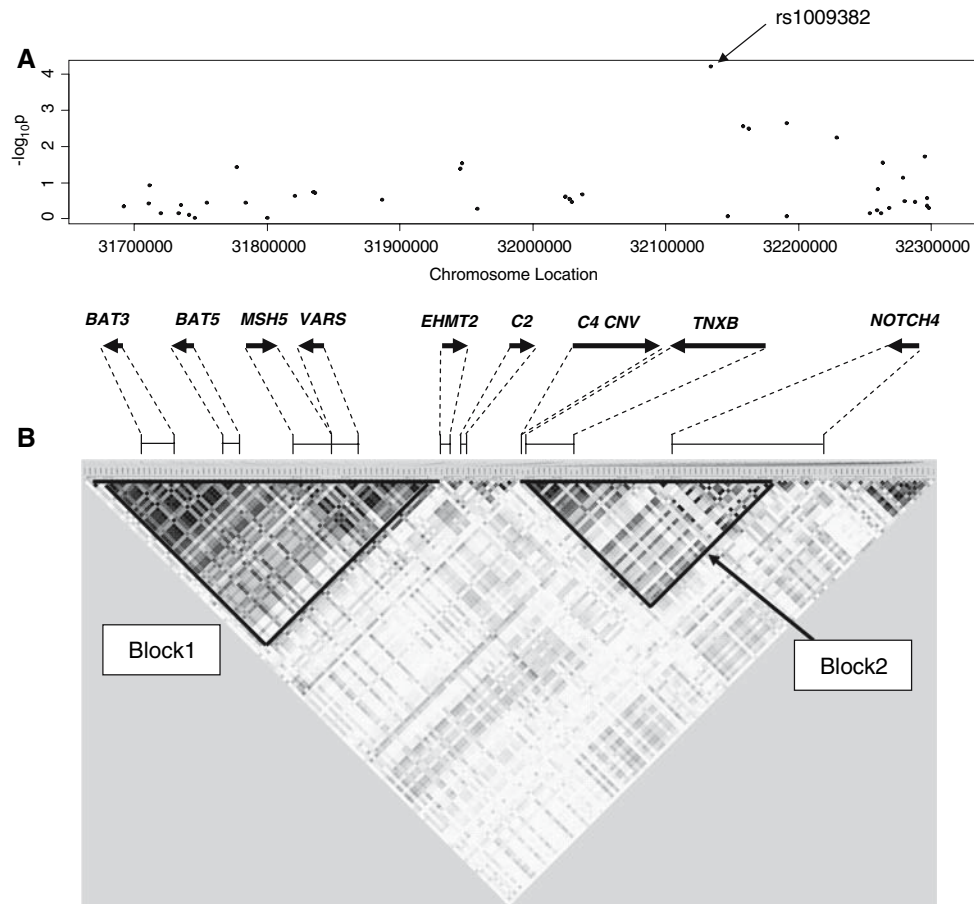


Fig. 1 Association of SNPs in the *TNXB* gene with SLE. **a** Single-point associations between SLE and tag SNPs around the *C4* CNV region. The Y axis is in units of $-\log_{10}(P)$. The lowest P value was detected at an SNP rs1009382 locus. **b** An LD map of this region

constructed from the HapMap-JPT data. The r^2 value between a pair of SNPs (minor allele frequency was 0.2 or more) was plotted at the apex of an *isosceles right triangle* defined by the baseline connecting the SNPs. Two LD blocks are shown by *bold edged triangle*

Systemic lupus erythematosus is known to occur predominantly in female patients; the female/male ratio in our study population was approximately 9:1. Since the control 2 population was not gender-matched, we performed a subgroup analysis using female patients and controls. The association study of SNP rs3130342 in 150 female patients and 67 female control subjects revealed a P value of 0.00395, in a recessive model, by Fisher’s exact test (Table 2). The G-allele frequency and the odds ratio were almost the same as when we used all case and control subjects. To validate further the association of rs3130342 with SLE susceptibility, we performed a replication study using independent cases (case 3, $n = 203$) and gender-matched controls (control 3, $n = 294$). As a result, the association of rs3130342 was replicated in a recessive-effect model ($P = 0.044$ at $\alpha = 0.05$, Table 2).

Among the many genes that were included in the human MHC region on chromosome 6p21, particular HLA alleles, as well as deletion of the *C4* gene, were suggested to be

genetic factors susceptible to SLE. The SNP rs3130342 is located close to the *C4* gene, and a recent study of a Caucasian SLE population indicated the association of a lower *C4* gene copy number with SLE susceptibility (Yang et al. 2007). Therefore, it is possible that the association between the SNP rs3130342 and SLE might simply reflect the linkage disequilibrium between *C4* CNV and SNP on *TNXB*. Hence, we investigated the number of copies of the *C4* gene and each of its isoforms (*C4A*, *C4B*, *C4L*, and *C4S*), in the DNA of 178 SLE patients (cases 1 and 2) and 365 control subjects (control 2), by qPCR.

On the basis of the results shown in Table 3, we tested the null hypothesis that the distribution of the copy number in the SLE patients was the same as in the healthy controls, using Wilcoxon’s rank sum test. The null hypothesis was rejected at significance level $\alpha = 0.01$ (determined by Bonferroni’s correction for five tests) in the analysis of copy numbers of total *C4* and *C4B* (P values of 0.0003453 and 0.0008674, respectively). Copy numbers of total *C4*

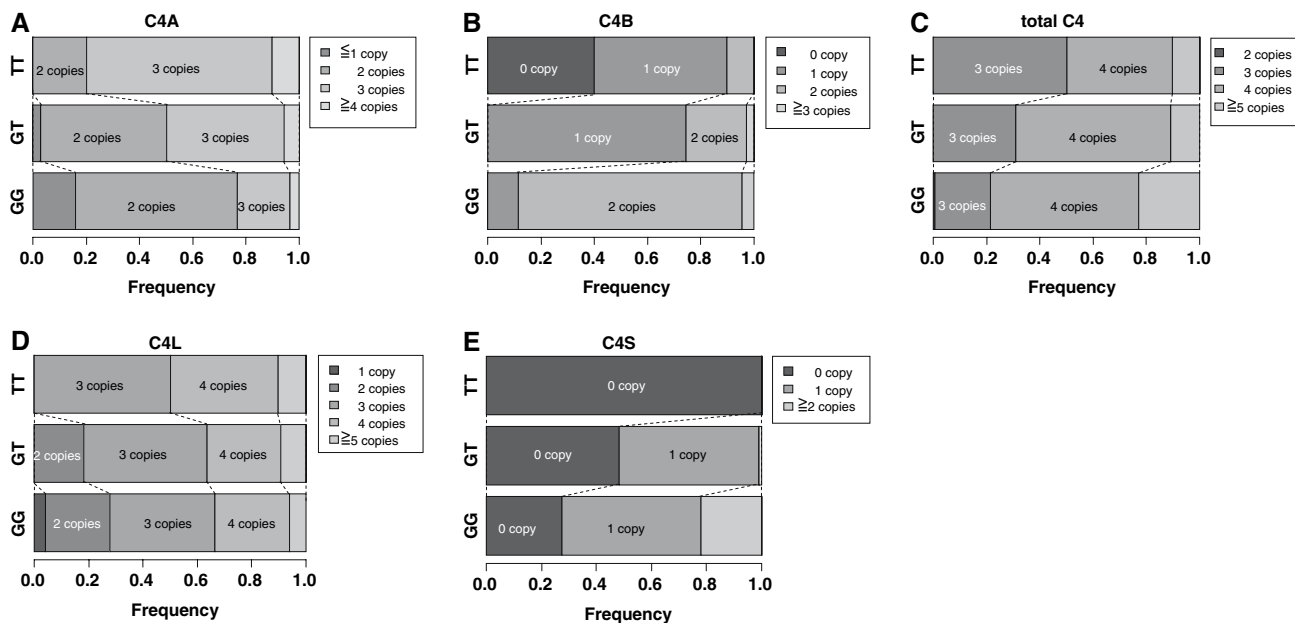


Fig. 2 Distributions of copy numbers of *C4* and its isoforms in the three genotype groups for an SNP rs3130342. The numbers of subjects were as follows: GG ($n = 244$), GT ($n = 109$), and TT ($n = 10$) at SNP rs3130342. **a** For *C4A*, two copies were prevalent in the GG genotype group, two and three copies were almost equal in the GT genotype group, and three copies were prevalent in the TT genotype group. **b** For *C4B*, more than 80% of the subjects with the GG genotype had two copies, and those with the GT genotype predominantly had one copy. The loss of the *C4B* gene was specifically observed in the group with the TT genotype. **c** Total

number of copies of the *C4* gene. Five or more copies were observed more frequently in the group with the GG genotype, and three copies were predominantly detected in the group with the TT genotype. **d** For *C4L*, none of the individuals with the TT genotype was judged to have two copies. **e** For *C4S*, the most significant difference was observed. All individuals belonging to the group with the TT genotype were judged to have no copy of *C4S*, while the remaining two-genotype groups predominantly possessed one or more copies. The GG genotype at SNP rs3130342 is a risk genotype for SLE

and *C4B* were significantly higher in SLE patients (4.20 and 1.88 in cases, and 3.96 and 1.74 in controls, respectively). Similar results were also obtained when we used the female subgroup for the analysis (Supplementary Table 4). Thus, our analyses indicated that greater numbers of copies of the *C4* and *C4B* gene might possibly be risk factors for SLE, although these results were contradictory to the results reported by Yang et al. (2007), in which lower numbers of copies of the *C4* gene were shown to be a susceptible factor to SLE.

Since the association of both *C4* CNV and SNP rs3130342 with SLE might simply reflect linkage disequilibrium between these two loci, we analyzed the relation between the number of copies of *C4* and its isoforms, and the genotypes of this particular SNP (Fig. 2). A Kruskal–Wallis test of the results revealed that copy numbers of *C4A* ($P = 4.213 \times 10^{-7}$, $\alpha = 0.01$ for all these five tests), *C4B* ($P = 1.195 \times 10^{-31}$), and *C4S* ($P = 7.32 \times 10^{-10}$) were significantly different among the three genotypes of the SNP, although the total *C4* copy number was of borderline significance ($P = 0.03734$). The number of copies of *C4A* became smaller according to the increase in the number of the G alleles (risk allele), while

the numbers of copies of *C4B*, total *C4*, and *C4S* increased when the number of G alleles was higher. These data suggested that SNP rs3130342 is in linkage disequilibrium with the CNV, especially with the copy number of *C4B*. The same tendency was also observed when we examined case 1 and case 2 individuals (data not shown).

We performed multivariate logistic regression analysis to evaluate the independence of these associated variants (Table 4). The SNP rs3130342 continued to show an association with SLE, regardless of the *C4* gene copy number, and revealed the smallest P value and the highest OR ($P = 0.0000954$, OR 3.24, 95%CI 1.80–5.87), indicating that the *TNXB* gene, but not the *C4* CNV, is the possible susceptible locus to SLE.

Particular HLA alleles were suggested to be genetic factors susceptible to SLE in the MHC region. Among them, HLA-DRB1*1501 has repeatedly been reported to be associated with SLE in the Japanese population (Hashimoto et al. 1994; Ohashi et al. 2001). To exclude the possibility that the association of SNP rs3130342 with SLE was due to its linkage disequilibrium with this HLA locus, we genotyped HLA-DRB1 using the 178 patients with SLE and the 365 control subjects (cases 1

Table 1 Summary of association of SNPs on *TNXB* gene with SLE

Location	dbSNP	Chromosome Location	Allele ½	Case			Control			Allele 1 frequency		Minimum <i>P</i> value
				Genotype			Genotype			Case	Control	
				11	12	22	11	12	22			
5'flanking	–	32189783	T/G	167	7	1	336	28	1	0.97	0.96	0.2
5'flanking	rs13196724	32189400	G/C	0	4	170	0	9	356	0.01	0.01	1
5'flanking	rs9469084	32188361	G/A	133	42	1	299	60	5	0.88	0.9	0.085
5'flanking	rs3130342	32188124	G/T	152	23	1	244	110	10	0.93	0.82	9.3E-07
5'flanking	–	32187936	C/T	178	0	0	365	0	0	1	1	1
5'flanking	–	32187545	G/T	120	49	9	277	81	7	0.81	0.87	0.014
5'flanking	rs3807039	32186351	A/C	127	46	4	285	76	3	0.85	0.89	0.078
5'flanking	rs411377	32185357	A/G	4	36	122	5	60	298	0.14	0.1	0.068
Intron 1	rs2269426	32184476	T/C	25	77	75	48	163	153	0.36	0.36	0.79
Exon 3	–	32173091	G/A	142	35	1	307	46	5	0.9	0.92	0.083
Exon 3	rs41270461	32173001	G/A	178	1	0	358	7	0	1	0.99	0.28
Exon 3	rs204896	32172075	A/G	2	34	122	5	59	300	0.12	0.09	0.18
Exon 3	rs41270458	32171874	C/T	174	1	0	363	2	0	1	1	1
Exon 3	–	32171771	A/G	150	25	1	294	71	0	0.92	0.9	0.19
Exon 3	rs17201602	32171659	G/A	0	3	171	0	9	356	0.01	0.01	0.76
Intron 3	rs2071293	32170665	T/C	18	74	80	35	156	171	0.32	0.31	0.76
Intron 3	rs204899	32165605	A/G	2	40	132	5	60	295	0.13	0.1	0.11
Intron 3	–	32165126	G/A	165	6	1	334	29	1	0.98	0.96	0.099
Exon 5	rs204900	32164557	T/G	126	42	1	298	60	6	0.87	0.9	0.064
Intron 7	rs2239688	32162189	C/G	10	47	111	7	85	273	0.2	0.14	0.0085
Exon 10	rs185819	32158044	A/G	18	50	90	48	163	151	0.27	0.36	0.0016
Intron 11	rs204877	32155414	C/G	1	42	134	5	61	299	0.12	0.1	0.11
Intron 11	rs2269427	32155254	G/C	0	4	172	1	9	355	0.01	0.02	0.78
Intron 14	rs2071295	32146677	A/G	16	74	82	34	160	171	0.31	0.31	0.93
Exon 18	rs3749963	32144427	A/G	1	26	150	0	70	295	0.08	0.1	0.28
Intron 20	rs9469080	32143405	A/G	1	41	135	5	60	300	0.12	0.1	0.11
Intron 20	rs2239689	32138262	C/T	18	78	80	36	160	169	0.32	0.32	0.84
Exon 21	rs2269428	32138177	A/C	1	26	149	0	69	295	0.08	0.09	0.33
Exon 23	rs1009382	32134084	A/G	109	61	7	158	170	34	0.79	0.67	0.000062
Intron 25	–	32131521	C/T	161	16	0	335	27	1	0.95	0.96	0.62
Intron 26	rs11969759	32129108	G/A	148	26	1	319	43	2	0.92	0.94	0.35
Intron 29	–	32125013	G/A	178	0	0	365	0	0	1	1	1
Intron 30	–	32123842	G/A	4	18	153	6	58	300	0.07	0.1	0.17
Intron 30	–	32123774	T/C	16	21	138	10	65	287	0.15	0.12	0.0022

and 2 and control 2). Our data indicated an association of HLA-DRB1*1501 with SLE (allele frequency was 15.4% and 7.0% in case and control samples, respectively, $P = 0.0000245$), similar to the previous reports. Stratified analysis revealed strong association between SNP rs3130342 and SLE in subjects without the HLA-DRB1*1501 allele ($P = 0.0000727$ at recessive model, Supplementary Table 5). Thus, we concluded that the GG genotype of SNP rs3130342 was associated with SLE, independent of the HLA-DRB1*1501 allele.

Discussion

In this study, we found a significant association of the SNP rs3130342 in the *TNXB* gene with SLE susceptibility in the Japanese population. The *TNXB* gene is reported to encode a member of the tenascin family of extracellular matrix proteins that was highly expressed in skin, muscle, tendon sheath, peripheral nerve, and vasculature (Matsumoto et al. 1994; Geffrotin et al. 1995). Molecular pathogenesis of SLE was considered as the exaggerated B cell response

Table 2 Association of SNP rs3130342 with SLE

Parameter	Number	Case			Number	Control			Frequency of G allele		Odds ratio ^a (95%CI)	P value ^b
		Genotype				Genotype			Case	Control		
		GG	GT	TT		GG	GT	TT				
Case 1, 2/control 2	177	152	23	1	363	244	110	10	0.929	0.822	3.11 (1.92–5.05)	9.30×10^{-7}
Female ^c	150	130	19	1	66	45	17	3	0.93	0.822	2.87 (1.45–5.86)	0.00395
Case 3/control 3	203	155	44	4	294	200	90	4	0.872	0.833	1.52 (1.01–2.28)	0.044
Total ^d	380	306	67	5	657	444	200	14	0.888	0.827	2.05 (1.51–2.78)	2.50×10^{-6}

^a Odds ratio of recessive effect (GG or GT + TT)

^b Fisher's exact test of recessive effect model

^c Analysis using female subgroup from case 1, case 2 and control 1

^d Combined analysis comparing the sum of case 1, case 2 and case 3 with the sum of control 2 and control 3

Table 3 Association of gene copy number of C4 and its isoforms with SLE

C4A	0	1	2	3	4	5		Mean ± SD	P value ^a
Control 2	1	41	203	105	12	3		2.26 ± 0.74	
Case 1, Case 2	2	9	102	42	13	0		2.32 ± 0.75	0.3784
C4B	0	1	2	3	4		Mean ± SD	P value ^a	
Control 2	4	115	232	13	1		1.70 ± 0.56		
Case 1, Case 2	3	26	129	9	1		1.88 ± 0.54	0.0003453*	
C4	2	3	4	5	6	7	Mean ± SD	P value ^a	
Control 2	1	90	206	60	5	3	3.96 ± 0.74		
Case 1, Case 2	1	18	107	31	10	1	4.20 ± 0.76	0.0008674*	
C4L	1	2	3	4	5	6	7	Mean ± SD	P value ^a
Control 2	10	78	149	102	22	2	2	3.17 ± 0.98	
Case 1, Case 2	0	33	67	56	7	4	1	3.32 ± 0.95	0.1536
C4S	0	1	2	3			Mean ± SD	P value ^a	
Control 2	131	179	54	1			0.79 ± 0.69		
Case 1, Case 2	48	91	29	0			0.89 ± 0.67	0.1225	

^a Results of Wilcoxon rank sum test comparing cases 1 + 2 with control 2. The significance level alpha is 0.01

*P values below significance level

Table 4 Results of multivariate logistic regression analysis

Variables	P value	OR	95% CI
SNP rs3130342 ^a	0.0000954	3.24	1.80–5.87
C4A CNV ^b	0.0137	1.41	1.08–1.84
C4B CNV ^b	0.223	1.33	0.86–2.06

^a Value 1 for GG genotype and value 0 for GT or TT genotypes

^b Values are the numbers of copies

caused by activated T cells or dendritic cells, or soluble mediators such as cytokines (Kyttaris et al. 2005). Tissue deposition of antibodies or immune complex induces

inflammation and subsequent injury of multiple organs. Mutations in the *TNXB* gene were found in patients with Ehlers–Danlos-like syndrome (Burch et al. 1997), but the role of *TNXB* in the immune system was not known.

C4 has an essential role in the classical complement pathway. C4A deficiency was repeatedly reported to be associated with SLE in Caucasian populations. However, the association of C4 CNV with SLE was not examined until quite recently. In 2002, the genomic structure of the C4 CNV region was precisely determined (Chung et al. 2002), and the association of the lower copy number of the C4 or C4A gene with SLE in the Caucasian population was reported recently (Yang et al. 2007). However, our result

that a higher number of copies of the *C4* gene was associated with the susceptibility to SLE was controversial with regard to their results. In earlier days, the frequency of an 8.5 kb-*HindIII* band (corresponding to *C4A* gene deletion) was reported to be higher in SLE patients (20.5–24.5%) than in control individuals (7.9–12.9%) in the Caucasian population (Kemp et al. 1987; Goldstein et al. 1988; Reveille et al. 1991; Hartung et al. 1992), while in an Afro-American population, its frequencies were 14.5% in SLE patients and 3.7% in controls (Olsen et al. 1989). On the other hand, this 8.5 kb-*HindIII* band itself was not observed in Japanese, Korean, or Chinese populations (Yamada et al. 1990; Doherty et al. 1992; Hong et al. 1994), indicating that this particular allele was extremely rare or absent in Asian populations. Two basepair insertion mutations that would cause a premature stop codon in exon 30 of the *C4A* gene were reported in Caucasian SLE patients (Barba et al. 1993; Lokki et al. 1999), but we found such mutation in none of our SLE patients (data not shown).

Considering the ethnic differences indicated above, the discrepancy between our result (association of a higher copy number of *C4* to SLE susceptibility) and the result by Yang et al. (association of a lower copy number of *C4* to SLE susceptibility) could be explained by historical recombination, where the true susceptible risk allele is linked to the high copy number of *C4* in Japanese populations and with the low copy number in Caucasians. In general, the existence of linkage disequilibrium between SNP locus and CNV locus is a matter of argument (Locke et al. 2006; Redon et al. 2006), but, especially in this region, we clearly indicated that there was linkage disequilibrium between *C4* CNV locus and the SNP on *TNXB*. On multivariate logistic regression analysis, the SNP rs3130342 remained associated with SLE, independent of *C4* CNV, and this result suggests that the association between *C4* CNV and SLE might only reflect the linkage disequilibrium between *C4* CNV and SNP on *TNXB*.

We also examined the effect of HLA locus on SNP rs3130342. Since HLA-DRB1*1501 was shown to associate with SLE in Japanese populations, we examined the HLA-DRB1 gene and revealed that SNP rs3130342 was still associated with SLE in subjects without the HLA-DRB1*1501 allele. We also found the tendency of association ($P = 0.092$, odds ratio 2.94, 95%CI 0.76–13.6) in individual with the HLA-DRB1*1501 allele, although the association was not statistically significant, possibly due to smaller sample size ($n = 52$ and 51 in case and control samples). HLA-B*0801-HLA-DRB1*0301 haplotype was also frequently reported to be associated with SLE in Caucasians. This haplotype was included in a large haplotype called 8.1 ancestral haplotype or “COX haplotype” that consists of HLA-A1 Cw7 B8 C4A Q0 C4B1

DRB1*0301 DRB3*0101 DQB1*0201 DQA1*0501 (Stewart et al. 2004). Moreover, the risk allele (G allele) of SNP rs3130342 was included in the COX haplotypes in Caucasian populations according to the Sanger MHC haplotype project. However, the HLA-DRB1*0301 allele was almost absent or extremely rare in the Japanese population (Hashimoto et al. 1994) and none carried the HLA-DRB1*0301 allele in our case and control individuals. Taken together, the association of rs3130342 with SLE was not due to its linkage disequilibrium with “COX haplotype” or HLA-DRB1*1501 in Japanese populations, and rs3130342 might have a causative role in SLE pathogenesis.

In summary, we performed a genome-wide association study of SLE and identified the SNP rs3130342 in the 5' flanking region of the *TNXB* gene as a possible candidate gene susceptible to the disease. The *TNXB* gene encodes an extracellular matrix protein, tenascin XB, that regulates collagen synthesis and deposition (Mao et al. 2002). Our findings indicated the possible role of collagen metabolism in the pathogenesis of SLE, and the association of *C4* CNV might not have a causal effect in the Japanese population. However, to clarify the physiological role of the *TNXB* polymorphism in SLE susceptibility, further investigation are required.

Web resources

Pharma SNP Consortium, <http://www.jpma.or.jp/psc/index.html>. The R statistical software: <http://www.R-project.org>. The HapMap project: <http://www.hapmap.org>. The JSNP database: <http://snp.ims.u-tokyo.ac.jp/index.html>. NCBI BLAST: <http://www.ncbi.nlm.nih.gov/BLAST>. The Sanger MHC haplotype project: <http://www.sanger.ac.uk/HGP/Chr6/MHC/>.

Acknowledgments We are grateful to members of The Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study.

References

- Barba G, Rittner C, Schneider PM (1993) Genetic basis of human complement C4A deficiency. Detection of a point mutation leading to nonexpression. *J Clin Invest* 91:1681–1686
- Barrett JC, Fry B, Maller J, Daly MJ (2004) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Burch GH, Gong Y, Liu W, Dettman RW, Curry CJ, Smith L, Miller WL, Bristow J (1997) Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet* 17:104–108
- Chung EK, Yang Y, Rennebohm RM, Lokki ML, Higgins GC, Jones KN, Zhou B, Blanchong CA, Yu CY (2002) Genetic sophistication of human complement components C4A and C4B and

- RP-C4-CYP21-TNX (RCCX) modules in the major histocompatibility complex. *Am J Hum Genet* 71:823–837
- Croker JA, Kimberly RP (2005) Genetics of susceptibility and severity in systemic lupus erythematosus. *Curr Opin Rheumatol* 17:529–537
- Dangel AW, Baker BJ, Mendoza AR, Yu CY (1995) Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K (C4) are a molecular clock of evolution. *Immunogenetics* 42:41–52
- Doherty DG, Ireland R, Demaine AG, Wang F, Veerapan K, Welsh KI, Vergani D (1992) Major histocompatibility complex genes and susceptibility to systemic lupus erythematosus in southern Chinese. *Arthritis Rheum* 35:641–646
- Forabosco P, Gorman JD, Cleveland C, Kelly JA, Fisher SA, Ortman WA, Johansson C, Johanneson B, Moser KL, Gaffney PM, Tsao BP, Cantor RM, Alarcon-Riquelme ME, Behrens TW, Harley JB, Lewis CM, Criswell LA (2006) Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. *Genes Immun* 7:609–614
- Geffrotin C, Garrido JJ, Tremet L, Vaiman M (1995) Distinct tissue distribution in pigs of tenascin-X and tenascin-C transcripts. *Eur J Biochem* 231:83–92
- Goldstein R, Arnett FC, McLean RH, Bias WB, Duvic M (1988) Molecular heterogeneity of complement component C4-null and 21-hydroxylase genes in systemic lupus erythematosus. *Arthritis Rheum* 31:736–744
- Hartung K, Baur MP, Coldewey R, Fricke M, Kalden JR, Lakomek HJ, Peter HH, Schendel D, Schneider PM, Seuchter SA, Stangel W, Deicher HRG (1992) Major histocompatibility complex haplotypes and complement C4 alleles in systemic lupus erythematosus. Results of a multicenter study. *J Clin Invest* 90:1346–1351
- Hashimoto H, Nishimura Y, Dong RP, Kimura A, Sasazuki T, Yamanaka K, Tokano Y, Murashima A, Kabasawa K, Hirose S (1994) HLA antigens in Japanese patients with systemic lupus erythematosus. *Scand J Rheumatol* 23:191–196
- Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 40:1725
- Hong GH, Kim HY, Takeuchi F, Nakano K, Yamada H, Matsuta K, Han H, Tokunaga K, Ito K, Park KS (1994) Association of complement C4 and HLA-DR alleles with systemic lupus erythematosus in Koreans. *J Rheumatol* 21:442–447
- Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, Kamatani N (2006) Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology* 45:1345–1348
- Kawai S, Maekawajiri S, Tokunaga K, Kashiwase K, Miyamoto M, Akaza T, Juji T, Yamane A. (1996) Routine low and high resolution typing of the HLA-DRB gene using the PCR-MPH (microtitre plate hybridization) method. *Eur J Immunogenet* 23:471–486
- Kemp ME, Atkinson JP, Skanes VM, Levine RP, Chaplin DD (1987) Deletion of C4A genes in patients with systemic lupus erythematosus. *Arthritis Rheum* 30:1015–1022
- Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Sueishi K, Iida M, Nakamura Y, Kiyohara Y (2007) A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet* 39:212–217
- Kyttaris VC, Juang YT, Tsokos GC (2005) Immune cells and cytokines in systemic lupus erythematosus: an update. *Curr Opin Rheumatol* 17:518–522
- Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* 79:275–290
- Lokki ML, Circolo A, Ahokas P, Rupert KL, Yu CY, Colten HR (1999) Deficiency of human complement protein C4 Due to identical frameshift mutations in the C4A and C4B genes^{1,2}. *J Immunol* 162:3687–3693
- Mao JR, Taylor G, Dean WB, Wagner DR, Afzal V, Lotz JC, Rubin EM, Bristow J (2002) Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. *Nat Genet* 30:421–425
- Matsumoto K, Saga Y, Ikemura T, Sakakura T, Chiquet-Ehrismann R (1994) The distribution of tenascin-X is distinct and often reciprocal to that of tenascin-C. *J Cell Biol* 125:483–493
- Ohashi J, Yamamoto S, Tsuchiya N, Hatta Y, Komata T, Matsushita M, Tokunaga K (2001) Comparison of statistical power between 2* 2 allele frequency and allele positivity tables in case-control studies of complex disease genes. *Ann Hum Genet* 65:197–206
- Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y (2001) A high-throughput SNP typing system for genome-wide association studies. *J Hum Genet* 46:471–477
- Olsen ML, Goldstein R, Arnett FC, Duvic M, Pollack M, Reveille JD (1989) C4A gene deletion and HLA associations in black Americans with systemic lupus erythematosus. *Immunogenetics* 30:27–33
- Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 32:650–654
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shaper MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. *Nature* 444:444–454
- Reveille JD, Anderson KL, Schrohenloher RE, Acton RT, Barger BO (1991) Restriction fragment length polymorphism analysis of HLA-DR, DQ, DP and C4 alleles in Caucasians with systemic lupus erythematosus. *J Rheumatol* 18:14–18
- Saito A, Kamatani N (2002) Strategies for genome-wide association studies: optimization of study designs by the stepwise focusing method. *J Hum Genet* 47:360–365
- Seki S, Kawaguchi Y, Chiba K, Mikami Y, Kizawa H, Oya T, Mio F, Mori M, Miyamoto Y, Masuda I, Tsunoda T, Kamata M, Kubo T, Toyama Y, Kimura T, Nakamura Y, Ikegawa S (2005) A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. *Nat Genet* 37:607–612
- Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, Forbes S, Halls K, Howson JM, Humphray SJ, Hunt S, Mungall AJ, Osoegawa K, Palmer S, Roberts AN, Rogers J, Sims S, Wang Y, Wilming LG, Elliott JF, Jong PJ, Sawcer S, Todd JA, Trowsdale J, Beck S (2004) Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 14:1176–1187
- Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, Nagasaki M, Nakayama-Hamada M, Kawaida R, Ono M, Ohtsuki M, Furukawa H, Yoshino S, Yukioka M, Tohma S, Matsubara T, Wakitani S, Teshima R, Nishioka Y, Sekine A, Iida A, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K (2003) Functional haplotypes of PADI 4, encoding citrullinating

- enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 34:395–402
- Szilagy A, Blasko B, Szilassy D, Fust G, Sasvari-Szekely M, Ronai Z (2006) Real-time PCR quantification of human complement C 4 A and C 4 B genes. *BMC Genet* 7:1
- Trager J, Ward MM (2001) Mortality and causes of death in systemic lupus erythematosus. *Curr Opin Rheumatol* 13:345–351
- Tsao BP (2003) The genetics of human systemic lupus erythematosus. *Trends Immunol* 24:595–602
- Tsunoda T, Lathrop GM, Sekine A, Yamada R, Takahashi A, Ohnishi Y, Tanaka T, Nakamura Y (2004) Variation of gene-based SNPs and linkage disequilibrium patterns in the human genome. *Hum Mol Genet* 13:1623–1632
- Yamada H, Watanabe A, Mimori A, Nakano K, Takeuchi F, Matsuta K, Tanimoto K, Miyamoto T, Yukiya Y, Tokunaga K, Yokohari R (1990) Lack of gene deletion for complement C4A deficiency in Japanese patients with systemic lupus erythematosus. *J Rheumatol* 17:1054–1057
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride KL, Higgins GC, Rennebohm RM, Rice RR, Hackshaw KV, Roubey RAS, Grossman JM, Tsao BP, Birmingham DJ, Rovin BH, Hebert LA, Yu CY (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037–1054