Association of *EMCN* with Susceptibility to Rheumatoid Arthritis in a Japanese Population

KAZUMASA NISHIMOTO, KATSUNORI IKARI, HIROTAKA KANEKO, SO TSUKAHARA, YUTA KOCHI, KAZUHIKO YAMAMOTO, YUSUKE NAKAMURA, YOSHIAKI TOYAMA, ATSUO TANIGUCHI, HISASHI YAMANAKA, and SHIGEKI MOMOHARA

ABSTRACT. ***Objective.*** Endomucin, an endothelial-specific sialomucin, is thought to facilitate "lymphocyte homing" to synovial tissues, resulting in the major histopathologies of rheumatoid arthritis (RA). We examined the association between RA susceptibility and the gene coding endomucin, *EMCN*.

Methods. Association studies were conducted with 2 DNA sample sets (initial set of 1504 patients, 752 controls; and validation set, 1113 patients, 940 controls) using 6 tag single-nucleotide polymorphisms (SNP) from the Japanese HapMap database. Immunohistochemistry for the expression of endomucin was conducted with synovial tissues from 4 patients with RA during total knee arthroplasty. Electromobility shift assays were performed for the functional study of identified polymorphisms.

Results. Within the initial sample set, the strongest evidence of an association with RA susceptibility was SNP rs3775369 (OR 1.20, p = 0.0075). While the subsequent replication study did not initially confirm the observed significant association (OR 1.13, p = 0.062), an in-depth stratified analysis revealed significant association in patients testing positive to anti-cyclic citrullinated peptide (anti-CCP) antibody in the replication data set (OR 1.15, p = 0.044). Investigating 2 sample sets, significant associations were detected in overall and stratified samples with anti-CCP antibody status (OR 1.17, p = 0.0015). Positive staining for endomucin was detected in all patients. The allele associated with RA susceptibility had a higher binding affinity for HEK298-derived nuclear factors compared to the nonsusceptible allelic variant of rs3775369.

Conclusion. A significant association between *EMCN* and RA susceptibility was detected in our Japanese study population. The *EMCN* allele conferring RA susceptibility may also contribute to the pathogenesis of RA. (First Release Dec 15 2010; J Rheumatol 2011;38:221–8; doi:10.3899/jrheum.100263)

Key Indexing Terms:
RHEUMATOID ARTHRITIS   GENE   SUSCEPTIBILITY   ENDOMUCIN

The contribution of genetic factors to the etiology of rheumatoid arthritis (RA) has been well established through extensive genealogical and twin-based studies[1,2]. HLA-DRB1 has been primarily associated with RA suscep-

tibility among most major ethnic groups, including those of European and Asian descent[3,4,5]. Many other genes are also believed to contribute to the pathogenesis of RA, presumably operating at small to modest risk levels.

Over the past decade, significant efforts have been made to identify putative RA susceptibility genes residing outside the HLA region. Often, these rely on a candidate gene or systematic approach, including genome-wide linkage scans. Such efforts have implicated *PTPN22* and *PADI4* as causative genes for RA susceptibility in European and Asian populations, respectively[6,7,8]. Additionally, *STAT4* has been shown to contribute a similar level of genetic risk for development of RA in European and Asian populations[9,10,11,12]. As well, *TNFAIP3*, *CD40*, *PRKCQ*, *KIF5A*, and *CCL21* were identified as RA-associated loci, by means of chip-based, genome-wide association studies and comprehensive replication studies among people of European origin[13,14,15,16,17,18].

However, more than half the genetic contribution to pathogenesis of RA is purported to arise from as-yet unknown genetic variance[18]. Considering that multiple molecular pathways are implicated in the onset and progression of the disease, it is likely that undetected genes of small to modest risk ultimately contribute to the pathogenesis of RA. Detection of such unknown genes outside the well studied HLA region is essential in understanding the genetic basis for RA susceptibility.

RA is characterized by chronic inflammation of the synovial tissues in joints and tendons. Entry of mononuclear cells, particularly T lymphocytes, into the synovium and the subsequent formation of aggregates into secondary ectopic lymphoid structures contributes to the systematic inflammation associated with RA via production of proinflammatory chemokines or pathogenic autoantibodies[19,20,21,22]. Recruitment of lymphocytes to synovial tissues is one of the key steps in the pathogenesis of RA. The homing of lymphocytes to secondary lymphoid structures is initiated by the interaction between sialomucin, a protein secreted on the endothelial cell surface, and cell adhesion molecules on lymphocytes, such as L-selectin. This group of several endothelial adhesion molecules that collectively bind L-selectin are commonly termed peripheral node addressins (PNAd). The secretion of PNAd in the vasculature of RA synovial tissues can be detected by immunostaining with the antiadhesive molecule MECA-79[20,23]. Thus, the PNAd are thought to play an important role in RA pathogenesis by facilitating lymphocyte homing to synovial tissues.

Endomucin, a PNAd, is known to be expressed in vascular endothelial cells of a variety of tissues including the heart, skin, and kidney[24,25]. While the expression of endomucin in synovial tissue has not been identified to date, it facilitates lymphocyte homing by interacting with L-selectin on respective lymphocytes like other PNAd[26]. Additionally, endomucin is overexpressed as a result of stimulation of tumor necrosis factor-α, a pleiotropic inflammatory cytokine central to a range of autoimmune diseases[27]. The implication of endomucin's involvement in key molecular processes suggests that *EMCN*, the endomucin coding gene, is a candidate susceptibility gene associated with RA. Our objective was to assess the association between *EMCN* and RA susceptibility within a Japanese population, and to investigate which, if any, associated polymorphisms have any directly identifiable biological effect on disease susceptibility.

MATERIALS AND METHODS

Study subjects. The study was approved by the relevant local ethics committee of each participating organization. Written informed consent was obtained from all participants prior to enrollment. All patients with RA had been diagnosed according to the revised criteria devised by the American College of Rheumatology[28].

Initial set. For initial analysis, case subjects were recruited from the Institute of Rheumatology Rheumatoid Arthritis (IORRA) associated with Tokyo Women's Medical University[29]. The IORRA series, including a case-control study for each haplotype tag single-nucleotide polymorphism (SNP) and haplotype analysis, consisted of 1504 cases and 752 controls. Control subjects were obtained from the DNA collection of the Pharma SNP Consortium, Tokyo, Japan, currently entrusted to the Health Science Research Resources Bank, Osaka, Japan.

Validation set. In addition, a RIKEN/BioBank series consisting of 1113 cases and 940 controls was prepared for a validation study of the SNP that had the best evidence of association with the initial set from IORRA. Case subjects were derived from the BioBank Japan Project DNA collection and controls were recruited from the Rotary Club of Osaka-Midosuji District 2600 Rotary International in Japan[29].

SNP selection and genotyping. Six SNP were selected that provided an adequate description of the haplotypes detected in the Japanese-derived genotype data obtained from HapMap (http://www.hapmap.org) using HaploView 4.1[30]. SNP genotyping data were extracted from *EMCN* in the region 3 kb upstream from the start of the first identified exon and 1 kb downstream from the end of the last exon, as described[31,32]. SNP of minor allele frequency (> 0.1) were included, while SNP deviating from Hardy-Weinberg equilibrium ($p < 0.05$) were excluded. A total of 33 SNP were identified and 6 tag SNP genotyped in this study revealed 32 SNP (96%). Genotyping was performed using the TaqMan SNP genotyping assay, following the manufacturer's specifications (Applied Biosystems, Tokyo, Japan)[7]. Duplicate samples and negative controls were included to ensure the accuracy of the genotyping protocol.

Stratified analyses on anti-CCP antibody status. Owing to the heterogeneity of RA, and the occurrence of distinct subsets characterized by the presence or absence of anticitrullinated peptide antibody (anti-CCP)[33], a stratified analysis was conducted for both the IORRA and RIKEN/BioBank-derived sample sets. The serum level of anti-CCP antibody was measured by a standard ELISA.

Statistical analysis. The differences in allele and genotype frequencies of each SNP between cases and controls were examined by the chi-squared test and the Cochran-Armitage test for trend, respectively. Chi-squared distribution was additionally applied to the stratified analyses on the level of anti-CCP antibody in RA patients, as well as in the statistical assessment of Hardy-Weinberg equilibrium. In subsequent combined analyses of the 2 case-controlled studies, a Breslow-Day test was used to assess significant heterogeneity and the Mantel-Haenszel procedure was performed to provide a common odds ratio when heterogeneity was denied. Statistical power for chi-squared test was calculated with alpha set as 0.05. These analyses were performed using the R software package (http://www.r-project.org/). Haplotypes were constructed using HaploView 4.1[30]. Multiple comparisons were corrected using the Bonferroni method. Biological interaction between the identified SNP and anti-CCP antibody status was evaluated with departure from additivity of effects. To quantify the amount of interaction, the attributable proportion due to interaction (AP) was calcu-

lated based on relative risks from logistic models using SPSS for Windows (version 17.0; SPSS Inc., Chicago, IL, USA). If there is no interaction, AP is equal to 0. A detailed description regarding definition and calculation of AP is described elsewhere[34,35].

Immunohistochemistry. Synovial tissues were obtained from 4 RA patients during total knee arthroplasty. Each specimen was frozen in liquid nitrogen within optimal cutting-temperature compound and cut on a cryostat. The expression of endomucin in derived synovial tissues was evaluated by immunohistochemical assays using the HISTAR detection system (STAR3000A; AbD Serotec, Kidlington, UK) in accord with manufacturer's specifications. Rat anti-human-endomucin antibody (ReproCell Inc., Yokohama, Japan) was used as a primary antibody and rat IgG was used as a negative control.

Electromobility shift assays (EMSA). HEK293 cell lines were cultured in Dulbecco's modified Eagle's medium containing fetal bovine serum albumin (10%), penicillin (1%), and streptomycin (1%). Nuclear extracts were prepared using an NE-PER Nuclear Extraction Reagent Kit (Pierce Biotechnology, Cramlington, UK) in accord with specifications. Nuclear extracts were stored in aliquots at –80°C until required. Biotin-labeled and unlabeled single-strand oligonucleotide probes (20 bp) were purchased (Invitrogen), corresponding to the haplotype-tagging SNP rs3775369 sequence of allele A (minus strand: 5'-GGT TAC ACC CAT GCG TAG CTG-3'; plus strand: 5'-CAG CTA CGC ATG GGT GTA ACC-3') and allele G (minus strand: 5'-GGT TAC ACC CGT GCG TAG CTG-3'; plus strand: 5'-CAG CTA CGC ACG GGT GTA ACC-3').

EMSA was performed using Lightshift chemiluminescent EMSA kit (Pierce). HEK293 nuclear extracts (2 μg) were incubated with the relevant biotin-labeled oligonucleotide probe for 20 min at 20°C in 10 mM Tris buffer (pH 7.5) containing 100 mM KCl, 1 mM DTT, 0.6 μg Poly(dI-dC), 0.05% NP-40, and 5 mM MgCl$_2$. For competition experiments, 25-, 50-, and 100-fold molar excesses of each unlabeled oligonucleotide probe were incubated 20 min with the nuclear extracts before addition of the labeled oligonucleotide probes. Each reaction mixture was then separated by electrophoresis on 6% polyacrylamide gel (Novex DNA Retardation Gel; Invitrogen) using 0.5 × TBE buffer. After electroblotting onto nylon membrane (Roche), the band shift was visualized according to the specifications of the Lightshift kit.

EMCN expression in relation to genotypes. EMCN expression data were extracted from the Genevar database (http://www.sanger.ac.uk/humgen/genevar/) for the HapMap samples as described[36]. The association between EMCN genotypes and expression was accessed using linear regression methods.

RESULTS

Association of SNP within EMCN locus with RA and haplotype analysis. All 6 of the identified SNP were in Hardy-Weinberg equilibrium within the identified disease cases and the controls. The genotyping success rate was over 99% for all SNP. Importantly, significant differences were observed in allele frequency between cases and controls in 4 of the 6 characterized SNP (Table 1). An SNP rs3775369 located within intron 1 showed the most significant association with disease susceptibility (OR 1.21, allele p = 0.0075, genotype p = 0.0070; Figure 1). After applying Bonferroni correction for multiple testing, only rs3775369 was still significant (allele p = 0.045, genotype p = 0.0042). The minor allele A in rs3775369 was associated with increased risk of RA.

Statistically, all 6 characterized SNP were in linkage disequilibrium. Four haplotypes were estimated to account for 90% of all haplotypes (Table 2). Significant differences between the identified disease cases and the controls were detected in the second and third common haplotype, haplotype 2 and haplotype 3, respectively. Haplotype 2 was more prevalent in RA patients (OR 1.19, p = 0.013); conversely, haplotype 3 was observed more frequently within the control group (OR 0.82, p = 0.036). However, these differences were no longer significant after Bonferroni correction (p = 0.09 and 0.29, respectively).

Validation study and combined analysis. To validate the association detected in the initial set, a replication study was conducted using an independent case-control study, derived from the RIKEN/BioBank sample set. The SNP rs3775369 was genotyped for the validation study as it demonstrated the strongest association. Although a recognizable trend of association was observed in the validation set, significant differences between the RA cases and controls were not detected (OR 1.13, allele p = 0.062, genotype p = 0.0058). Because no heterogeneity of allelic distribution was observed within the case-control sample set (IORRA and RIKEN/BioBank) (p = 0.54), all data were combined to provide common OR with significant statistical difference (Table 3; OR 1.17, allele p = 0.0015, genotype p = 0.0051).

Publicly available genome-wide association data with RA samples, all in Caucasian populations, were also analyzed to determine whether the association of EMCN with RA was supported[37]. We found no evidence of positive association between RA and the SNP located within the linkage disequilibrium block including EMCN in the database (data not shown).

Table 1. Case-control data on HapMap tag single-nucleotide polymorphisms of *EMCN* locus in the IORRA series.

| SNP | Location | Position* | Allele 1/2 | Genotypes of Patients | | | | | Genotypes of Controls | | | | | OR (95% CI) | p | |
				1/1	1/2	2/2	Total	MAF	1/1	1/2	2/2	Total	MAF		Allele	Genotype
rs3816056	Intron 1	25986294	G/T	1072	383	39	1494	0.154	495	236	20	751	0.184	0.81 (0.69–0.96)	0.012	0.012
rs3775369	Intron 1	25980974	G/A	681	669	145	1495	0.321	384	301	59	744	0.282	1.21 (1.05–1.38)	0.0075	0.0070
rs1501096	Intron 1	25980569	A/C	1037	405	46	1488	0.167	480	247	24	751	0.196	0.82 (0.70–0.97)	0.015	0.015
rs3796655	Intron 1	25980365	A/T	1043	407	46	1496	0.167	480	247	24	751	0.196	0.82 (0.70–0.96)	0.015	0.014
rs12501083	Intron 1	25970515	T/C	1218	265	14	1497	0.098	617	127	7	751	0.094	1.05 (0.84–1.30)	0.67	0.67
rs736517	Intron 2	25946759	G/C	1057	396	43	1496	0.161	511	217	24	752	0.176	0.90 (0.76–1.06)	0.20	0.20

* Positions are according to genomic contig NT_016354.18. MAF: minor allele frequency. IORRA: Institute of Rheumatology Rheumatoid Arthritis.
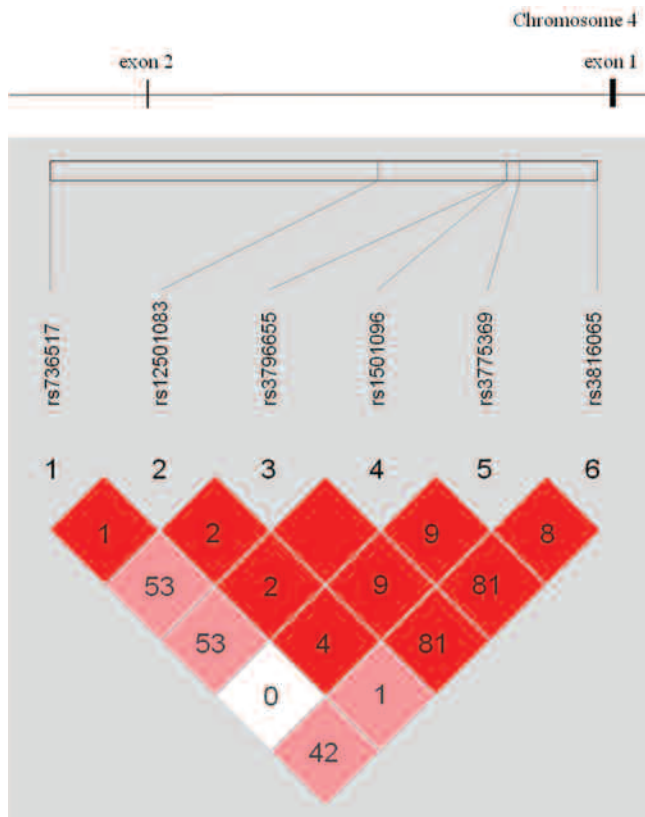
Figure 1. Linkage disequilibrium (LD) structure of 6 tagging SNP genotyped in all case and control subjects in the IORRA series. The key SNP rs3775369 is located within intron 1 of *EMCN*. Connected pairs of SNP show the standard color D'/LOD scheme using Haploview 4.1 (bright red: D' = 1, LOD score ≥ 2; shades of pink/red: D' < 1, LOD score ≥ 2; and white: D' < 1, LOD score < 2). Value in each diamond represents r^2.

Stratified analysis. Anti-CCP antibody measurements were available for the majority of patients in the IORRA data set (96%) and all patients (100%) in the RIKEN/BioBank data set. The frequency of the risk-inducing allele in patients returning positive results for the anti-CCP antibody was significantly higher than that in the controls, in both the IORRA and RIKEN/BioBank series (p = 0.012 and p = 0.044, respectively). Conversely, the incidence of patients with RA returning a negative result to the anti-CCP antibody was not significantly different from that in the control group in either of the sample series (Table 4). Combined analysis of both sample series also revealed a significant difference in the allelic distribution between RA patients who returned a positive result to the anti-CCP antibody and the control group (OR 1.17, p = 0.0015).

Interaction effects between rs3775369 and anti-CCP antibody status were estimated with a departure-from-additivity model by calculating the attributable proportion due to interaction in the IORRA data set. However, we found no evidence of a significant interaction (AP = 0).

Immunohistochemistry. Although endomucin is reportedly expressed in vascular endothelial cells of a variety of tissues, including the heart, skin, and kidney, expression in synovial tissues has not been elucidated. The expression of endomucin was thus examined by immunohistochemical staining in human synovial tissues from patients with confirmed RA following total knee arthroplasty. Within this sample set, positive staining confirmed the presence of endomucin in all patients, with expression specific to the endothelial cells of blood vessels (Figure 2). No staining was observed within the negative control (i.e., samples incubated without a primary antibody; data not shown).
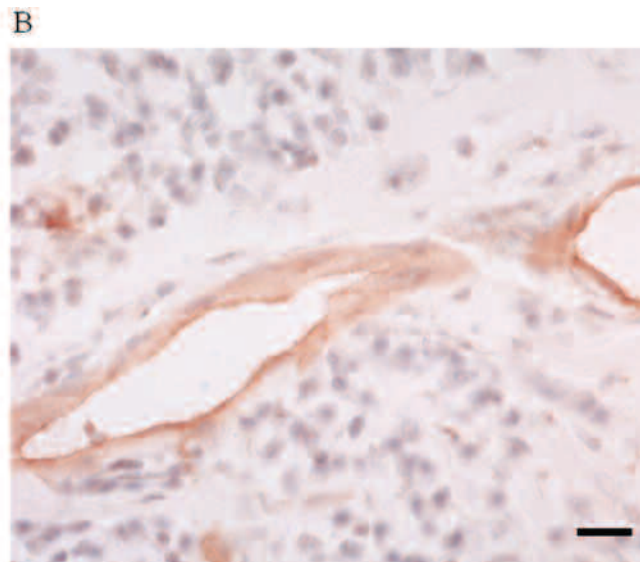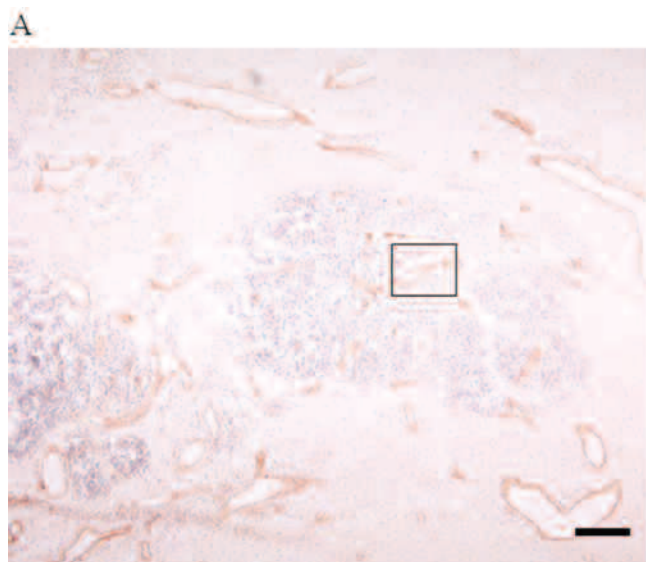

Figure 2. Immunohistochemical staining for expression of human endomucin in endothelial cells of vessels of RA synovial tissues. Brown stain indicates endomucin expression and sections are counterstained with hematoxylin (blue). A. Low-power image with inset shown in (B) at high-power magnification. Bars represent 200 μm (A) and 100 μm (B).

Table 2. EMCN haplotypes and association with RA.

Haplotype	a	b	c	d	e	f	Frequency Patients	Controls	OR (95% CI)	p	p_{corr}*
1	G	G	A	T	T	G	0.41	0.43	0.95 (0.84–1.08)	0.45	1.00
2	G	A	A	T	T	G	0.29	0.25	1.19 (1.03–1.38)	0.013	0.09
3	T	G	C	A	T	C	0.11	0.13	0.82 (0.67–0.99)	0.036	0.25
4	G	G	A	T	C	G	0.09	0.09	1.06 (0.85–1.32)	0.63	1.00
5	T	G	C	A	T	G	0.04	0.05	0.83 (0.61–1.14)	0.21	1.00
6	G	A	A	T	T	C	0.03	0.03	1.11 (0.77–1.63)	0.59	1.00
7	G	G	C	A	T	C	0.02	0.02	1.10 (0.68–1.84)	0.71	1.00

* After Bonferroni correction. a: rs3816056, b: rs3775369, c: rs1501096, d: rs3796655, e: rs12501083, f: rs736517.

Table 3. Validation study of association between rs3775369 and RA susceptibility in RIKEN/BioBank series and combined analysis.

Series	Genotypes of Patients GG	GA	AA	Total	MAF	Genotypes of Controls GG	GA	AA	Total	MAF	OR (95% CI)	p Allele	Genotype
RIKEN/BioBank	493	497	119	1109	0.331	442	416	76	934	0.304	1.13 (0.99–1.30)	0.062	0.058
IORRA	681	669	145	1495	0.321	384	301	59	744	0.282	1.21 (1.05–1.38)	0.0075	0.0070
Combined	1174	1166	264	2604	0.325	826	717	135	1678	0.294	1.17 (1.06–1.28)	0.0015	0.0051

MAF: minor allele frequency.

Table 4. Analyses of sr3775369 stratified on anti-CCP status in patients with RA.

Series	Status	No. Alleles G	A	Total	Frequency of Risk Allele (A)	Patients vs Controls OR (95% CI)	p
IORRA	Controls	1069	419	1488	0.28		
	Patients Anti-CCP (+)	1723	809	2532	0.32	1.20 (1.04–1.38)	0.012
	Anti-CCP (–)	248	120	368	0.33	1.23 (0.96–1.59)	0.092
RIKEN/BioBank	Controls	1300	568	1868	0.30		
	Patients Anti-CCP (+)	1225	617	1842	0.34	1.15 (1.00–1.33)	0.044
	Anti-CCP (–)	258	118	376	0.31	1.05 (0.82–1.34)	0.71
Combined	Patients Anti-CCP (+)	2948	1426	4374	0.33	1.17 (1.06–1.30)	0.0015
	Anti-CCP (–)	506	238	744	0.32	1.13 (0.96–1.35)	0.16

Anti-CCP: anti-cyclic citrullinated peptide antibody.

Electromobility shift assay (EMSA) results. Based on the finding of a significant association with susceptibility to RA, rs3775369 and other SNP identified in associated linkage disequilibrium may be the most likely candidates for causative SNP. However, a high level of linkage disequilibrium between rs3775369 and other SNP was not identified, according to the r^2 or LOD scores ($r^2 < 0.2$, LOD score < 2) derived from Haploview 4.1 in conjunction with the HapMap JPT database. The rs3775369 SNP was thus independently considered a likely candidate as a causative polymorphism. Owing to the location of rs3775369 within intron 1, allelic differences within this region may in turn alter the affinity for various nuclear factors, and hence the level of resultant transcriptions. To determine whether polymorphisms in rs3775369 affect the DNA binding abilities by various nuclear factors, an EMSA was conducted using nuclear extracts from human HEK293 cell lines and oligonucleotide probes corresponding to the sequence surrounding the SNP. Although labeled probes containing each alternative allele each produced retarded bands, significant differences in band intensity were not detected between the 2 recognized variants, the A and G alleles (Figure 3). Extending this technique to an additional competition analysis, excess non-labeled probe corresponding to the G allele did not significantly inhibit the ability of HEK293-derived nuclear extracts to bind to the labeled A allele. However, excess non-labeled probe corresponding to the A allele did in fact inhibit binding of nuclear extracts in a concentrate-dependent manner. This suggests a degree of difference in the binding ability of the 2 allelic variants, with allele A demonstrating higher binding affinity.

EMCN expression in relation to genotypes. Gene expression of EMCN was accessed with JPT HapMap subjects to evaluate whether there was any genotypic effect on EMCN
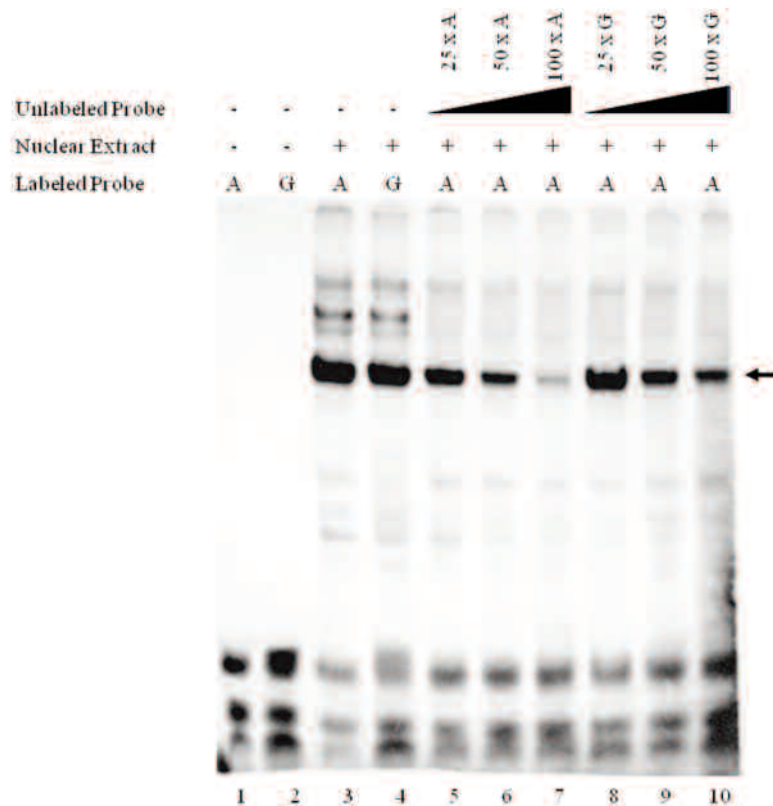
Figure 3. Electromobility shift assays with oligonucleotide probes corresponding to SNP rs3775369. Labeled probe alone containing allele A (lane 1) and allele G (lane 2); incubation with nuclear extracts from the HEK293 cell line, allele A (lane 3) and allele G (lane 4); competition experiments with increasing amounts of nonlabeled probe A (lanes 5–7) and nonlabeled allele G (lanes 8–10). Arrow indicates the band that was shifted with the addition of nuclear extracts.

expression. No association was observed between *EMCN* expression and rs3775369 genotypes (data not shown).

DISCUSSION
Our study identified a significant association between polymorphisms in *EMCN* and susceptibility to RA, using 2 independent sample sets within the Japanese population (OR 1.17, allele p = 0.0015, genotype p = 0.0051). The effect that *EMCN* itself has on disease susceptibility is moderate, based on the odds ratio obtained from this study. This finding is in accord with data for other RA susceptibility genes identified outside the HLA region; these also present a moderate risk. This is not surprising as it is generally accepted that the overall risk contribution of each genetic factor is small in common diseases such as RA. The complex etiology observed in RA likely results from multiple genetic and environmental factors, both known and unknown. To account for even a small level of associated risk, a large sample size was used to obtain sufficient statistical power. In our study, statistical power was calculated at 91% in the combined data set analysis.

The identified association between *EMCN* and RA sug-

gests that variation within *EMCN* may contribute to the development and subsequent pathogenesis of RA. Endomucin, as encoded by *EMCN*, is thought to regulate L-selectin-mediated lymphocyte homing on the surface of vascular endothelial cells[26]. We established that endomucin is indeed abundantly expressed in the vascular endothelial cells of RA synovial tissues. Endomucin could therefore serve to enhance inflammation of synovial tissues through the process of lymphocyte homing in RA patients. The genetic association between *EMCN* polymorphisms and RA identified in this study suggests that *EMCN* may contribute to the pathogenesis of the disease. Recently, a genetic association with CCL21, a chemokine also involved in lymphocyte homing, was identified in a powerful metaanalysis of 2 genome-wide association studies in Caucasians[18]. The recognition of additional genetic components that contribute to lymphocyte homing is critical for developing an understanding of the genetic basis of RA, and ultimately for identifying key targets for effective treatment of RA[38].

In stratified analyses derived from anti-CCP antibody assays, the association between *EMCN* and RA in both data sets was predominantly through the subgroup of RA patients

that tested positive to the autoantibody. However, this predominance might not be specific to this subgroup, since no interaction between *EMCN* and anti-CCP antibody status was found. Nevertheless, this correlation is perhaps understandable, as many studies have indicated that autoantibody-positive and negative RA may confer different levels of associated genetic risk[33]. Interestingly, ectopic lymphoid structures present in RA synovial tissues were reported to be involved in autoantibody production, including the anti-CCP antibody[19]. *EMCN* may indeed contribute to RA pathogenesis by facilitating lymphocyte homing to secondary lymphoid organs, subsequently causing the production of anti-CCP antibody.

While a positive association between *EMCN* and RA was detected in this study, there is still a possibility of false positives. False-positive associations may result from population stratification, or errors in sampling, phenotyping, genotyping, or multiple testing, as well as other potential biases[39,40]. Unaccounted or unidentified population stratification has the potential to produce spurious genetic associations. Although the Japanese population is known to have limited genetic diversity, 2 main population clusters have been identified, the Hondo and the Ryukyu[41]. To account for this, both case-control data series in our study were recruited from the Hondo area. Nevertheless, as the 2 data series were not tested for population stratification, this may remain a limitation and potential source of error.

Age and sex of cases and controls are not matched in the IORRA and RIKEN/BioBank series (IORRA cases, mean age 59.3 yrs, 83.8% female; controls, mean age 36.4 yrs, 50% female. RIKEN/BioBank cases, mean age 60.4 yrs, 82.0% female; controls, mean age 52.6 yrs, 25% female). This mismatch might also be a limitation of our study. With regard to errors derived from phenotyping, autoantibody testing reduced the possibility of phenotyping error by providing positive correlations with RA diagnosis. However, there remained some limited potential for phenotyping errors. Genotyping error was unlikely, owing to an overall genotyping success rate of 99% and a genotype concordance rate of 100%, assessed by duplicate samples. Although multiple testing does not eliminate false positives, the identified associations remained significant even after Bonferroni correction in the IORRA series (allele p = 0.045, genotype p = 0.0042).

The SNP allele frequencies of control subjects in this study had some deviations from JPT HapMap data, although the minor alleles of the SNP in this study were identical to those in the JPT HapMap data. This deviation might be a result of the relatively small numbers of samples used in the HapMap database. However, the possibility remains of false-positive results derived from the discrepancy.

Because the SNP rs3775369 is located within intron 1, it is suggested that alternative alleles may influence the binding ability of nuclear factors and thus their resultant transcription activity. Our EMSA revealed that the sequence containing the RA susceptibility allele A had a higher binding affinity to nuclear extracts than the alternative non-susceptible allele G. The Transcription Element Search System (TESS; http://www.cbil.upenn.edu/cgi-bin/tess/tess) predicted that the transcription factor AP-2α has the potential to bind to the sequence surrounding the SNP rs3775369 when the allelic variant is the RA susceptibility allele A. Variations in the *EMCN* allele may alter the binding affinity of AP-2α and hence regulate subsequent transcription of the gene. However, the genotypic effect on *EMCN* gene expression was not confirmed using the Genevar database with HapMap samples. Additional experiments are needed to fully deduce the changes of these genetic variation effects in the function of *EMCN*. Moreover, as the HapMap database does not fully account for the breadth of genetic variability, it is necessary to find other novel SNP that may also influence or determine the risk haplotype. Additional examination of the functional variants of *EMCN* is essential to establish a solid genetic association.

Our study clearly identified an association between *EMCN* and RA susceptibility within the Japanese population. The association was chiefly with RA patients who tested positive to anti-CCP antibody. This key finding suggests that variation within a gene engaged in lymphocyte homing may be critical in the pathogenesis of RA. The identified variation within the intronic SNP was significantly associated with RA and may alter transcription activity; further investigation is required to elucidate the mechanism of action. Identification of the RA disease-causing variant of *EMCN*, in conjunction with replicate studies in other populations, is central to reaching a definitive conclusion.

REFERENCES

1. Grant SF, Thorleifsson G, Frigge ML, Thorsteinsson J, Gunnlaugsdottir B, Geirsson AJ, et al. The inheritance of rheumatoid arthritis in Iceland. Arthritis Rheum 2001;44:2247-54.
2. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30-7.
3. Kochi Y, Yamada R, Kobayashi K, Takahashi A, Suzuki A, Sekine A, et al. Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. Arthritis Rheum 2004;50:63-71.
4. Louzada-Junior P, Freitas MV, Oliveira RD, Deghaide NH, Conde RA, Bertolo MB, et al. A majority of Brazilian patients with rheumatoid arthritis HLA-DRB1 alleles carry both the HLA-DRB1 shared epitope and anti-citrullinated peptide antibodies. Braz J Med Biol Res 2008;41:493-9.

5. Wordsworth BP, Lanchbury JS, Sakkas LI, Welsh KI, Panayi GS, Bell JI. HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. Proc Natl Acad Sci USA 1989;86:10049-53.

6. Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, et al. PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. Am J Hum Genet 2005;77:567-81.

7. Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, et al. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. Rheumatology 2006;45:1345-8.

8. Iwamoto T, Ikari K, Nakamura T, Kuwahara M, Toyama Y, Tomatsu T, et al. Association between PADI4 and rheumatoid arthritis: a meta-analysis. Rheumatology 2006;45:804-7.

9. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. Hum Mol Genet 2008;17:2274-9.

10. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977-86.

11. Kobayashi S, Ikari K, Kaneko H, Kochi Y, Yamamoto K, Shimane K, et al. Association of STAT4 with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population. Arthritis Rheum 2008;58:1940-6.

12. Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. Mol Med 2007;13:455-60.

13. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.

14. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nat Genet 2007;39:1431-3.

15. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477-82.

16. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet 2008;40:1156-9.

17. Orozco G, Eyre S, Hinks A, Ke X, Wilson AG, Bax DE, et al. Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study. Ann Rheum Dis 2010;69:813-6.

18. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 2008;40:1216-23.

19. Humby F, Bombardieri M, Manzo A, Kelly S, Blades MC, Kirkham B, et al. Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium. PLoS Med 2009;6:e1.

20. Pablos JL, Santiago B, Tsay D, Singer MS, Palao G, Galindo M, et al. A HEV-restricted sulfotransferase is expressed in rheumatoid arthritis synovium and is induced by lymphotoxin-alpha/beta and TNF-alpha in cultured endothelial cells. BMC Immunol 2005;6:6.

21. Takemura S, Braun A, Crowson C, Kurtin PJ, Cofield RH, O'Fallon WM, et al. Lymphoid neogenesis in rheumatoid synovitis. J Immunol 2001;167:1072-80.

22. Timmer TC, Baltus B, Vondenhoff M, Huizinga TW, Tak PP, Verweij CL, et al. Inflammation and ectopic lymphoid structures in rheumatoid arthritis synovial tissues dissected by genomics technology: identification of the interleukin-7 signaling pathway in tissues with lymphoid neogenesis. Arthritis Rheum 2007;56:2492-502.

23. Middleton J, Americh L, Gayon R, Julien D, Mansat M, Mansat P,

et al. A comparative study of endothelial cell markers expressed in chronically inflamed human tissues: MECA-79, Duffy antigen receptor for chemokines, von Willebrand factor, CD31, CD34, CD105 and CD146. J Pathol 2005;206:260-8.

24. Morgan SM, Samulowitz U, Darley L, Simmons DL, Vestweber D. Biochemical characterization and molecular cloning of a novel endothelial-specific sialomucin. Blood 1999;93:165-75.

25. Samulowitz U, Kuhn A, Brachtendorf G, Nawroth R, Braun A, Bankfalvi A, et al. Human endomucin: distribution pattern, expression on high endothelial venules, and decoration with the MECA-79 epitope. Am J Pathol 2002;160:1669-81.

26. Kanda H, Tanaka T, Matsumoto M, Umemoto E, Ebisuno Y, Kinoshita M, et al. Endomucin, a sialomucin expressed in high endothelial venules, supports L-selectin-mediated rolling. Int Immunol 2004;16:1265-74.

27. Liu C, Shao ZM, Zhang L, Beatty P, Sartippour M, Lane T, et al. Human endomucin is an endothelial marker. Biochem Biophys Res Commun 2001;288:129-36.

28. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.

29. Nishimoto K, Kochi Y, Ikari K, Yamamoto K, Suzuki A, Shimane K, et al. Association study of TRAF1-C5 polymorphisms with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in Japanese. Ann Rheum Dis 2010;69:368-73.

30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

31. Huang M, Han Y, Zhang X, Pei F, Deng J, Kang J, et al. An intron polymorphism in the CXCL16 gene is associated with increased risk of coronary artery disease in Chinese Han population: a large angiography-based study. Atherosclerosis;210:160-5.

32. Ong KL, Leung RY, Wong LY, Cherny SS, Sham PC, Lam TH, et al. Association of F11 receptor gene polymorphisms with central obesity and blood pressure. J Intern Med 2008;263:322-32.

33. Kurreeman FA, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, Stoeken-Rijsbergen G, et al. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med 2007;4:e278.

34. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol 2005;20:575-9.

35. Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. Am J Epidemiol 2003;158:251-8.

36. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Hikami K, et al. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum 2009;60:1845-50.

37. Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. BMC Med Genet 2009;10:6.

38. Yang J, Rosen SD, Bendele P, Hemmerich S. Induction of PNAd and N-acetylglucosamine 6-O-sulfotransferases 1 and 2 in mouse collagen-induced arthritis. BMC Immunol 2006;7:12.

39. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95-108.

40. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. Hum Hered 2007;64:203-13.

41. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: Effects on population-based association studies. Am J Hum Genet 2008;83:445-56.