nature
genetics

An intronic SNP in a RUNX1 binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis

Shinya Tokuhiro[1,2], Ryo Yamada[1], Xiaotian Chang[1], Akari Suzuki[1], Yuta Kochi[1,3], Tetsuji Sawada[3], Masakatsu Suzuki[2], Miyuki Nagasaki[2], Masahiko Ohtsuki[2], Mitsuru Ono[2], Hidehiko Furukawa[2], Masakazu Nagashima[4], Shinichi Yoshino[4], Akihiko Mabuchi[5], Akihiro Sekine[6], Susumu Saito[6], Atsushi Takahashi[7], Tatsuhiko Tsunoda[7], Yusuke Nakamura[8,9] & Kazuhiko Yamamoto[1,3]

Rheumatoid arthritis is a common inflammatory disease with complex genetic components. We investigated the genetic contribution of the cytokine gene cluster in chromosome 5q31 to susceptibility to rheumatoid arthritis in the Japanese population by case-control linkage disequilibrium (LD) mapping using single nucleotide polymorphisms (SNPs). Here we report that there is significant association between rheumatoid arthritis and the organic cation transporter gene *SLC22A4* (*P* = 0.000034). We show that expression of *SLC22A4* is specific to hematological and immunological tissues and that *SLC22A4* is also highly expressed in the inflammatory joints of mice with collagen-induced arthritis. A SNP affects the transcriptional efficiency of *SLC22A4 in vitro*, owing to an allelic difference in affinity to Runt-related transcription factor 1 (RUNX1), a transcriptional regulator in the hematopoietic system. A SNP in *RUNX1* is also strongly associated with rheumatoid arthritis (*P* = 0.00035). Our data indicate that the regulation of *SLC22A4* expression by RUNX1 is associated with susceptibility to rheumatoid arthritis, which may represent an example of an epistatic effect of two genes on this disorder.

Rheumatoid arthritis is a chronic inflammatory disease with autoimmune features that affects 0.5–1.0% of the world's population including the Japanese. Although the precise etiology of rheumatoid arthritis is unknown, genetic and environmental factors seem to be involved in its pathogenesis. The relative risk of developing the disorder is 2–17 times higher in siblings of affected individuals as compared to the general population, implying a genetic contribution to rheumatoid arthritis susceptibility[1]. Several genes are thought to contribute to susceptibility to rheumatoid arthritis[2–6].

As an alternative to linkage studies[4,6–8] and approaches based on target genes, LD mapping with SNPs for candidate regions coupled with whole-genome screening is considered to be a useful method for genetic association studies[9,10]. To identify genes associated with rheumatoid arthritis, a whole-genome LD mapping study using a high-throughput multiplex PCR-Invader assay[11,12] is currently in progress. Although incomplete, this project has previously identified the gene encoding peptidylarginine deiminase type IV as being associated with susceptibility to rheumatoid arthritis[13].

Inflammation of the synovium is one of the primary pathological features of rheumatoid arthritis. It is characterized by an increase in the number of inflammatory cells and the formation of granulation tissue with the subsequent destruction of joints. Autoimmunity is also considered to have an important role in the pathogenesis of rheumatoid arthritis because various autoantibodies are produced in conjunction with manifestations of rheumatoid arthritis[14]. The chromosomal region 5q31 is of particular interest in rheumatoid arthritis because it contains many genes involved in immune and inflammatory systems and also because it is suggested to be a susceptibility locus for several inflammatory or autoimmune diseases, including Crohn disease, atopic dermatitis and bronchial asthma, by studies on humans and mice[15–19]. For these reasons, we focused our search for genes associated with rheumatoid arthritis on the 5q31 cytokine cluster region using LD mapping with SNPs.

We report here that the gene encoding solute carrier family 22, member 4 (*SLC22A4*) on chromosome 5q31 is associated with rheumatoid arthritis. *SLC22A4* has been reported as encoding a transporter of organic molecules, and a region spanning 250 kb that

contains *SLC22A4* has been reported to be associated with Crohn disease, which, like rheumatoid arthritis, has a pathogenesis associated with inflammation and autoimmunity[16]. We show that *SLC22A4* is expressed in hematological tissues and cells and that a SNP located in a RUNX1-binding sequence in *SLC22A4* affects the expression of *SLC22A4* by altering RUNX1 binding affinity. RUNX1 is an essential hematopoietic transcription factor[20], mutations of which are responsible for acute myeloid leukemia[21]. A previous study has found that mutation of a RUNX1 binding site within *PDCD1* is associated with systemic lupus erythematosus[22]. Our findings indicate that *SLC22A4* may be involved in rheumatoid arthritis through the transportation of small organic molecules, and that polymorphisms in both *SLC22A4* and *RUNX1* are related to the expression of *SLC22A4* and thus to susceptibility to rheumatoid arthritis, thereby representing an example of multigenic contributions to a complex genetic trait.

RESULTS

LD evaluation and case-control association tests in 5q31

To study the 5q31 cytokine cluster region, we evaluated the whole region registered as NT_007072.11 on the National Center for

Biotechnology Information (NCBI) database. Initially, we evaluated the LD extension of the region with 172 SNPs registered in the Japanese Science and Technology Agency database[23] by genotyping 658 controls. For the LD block evaluation, we included SNPs with minor allele frequency >0.2, genotype success rate >0.9 and $P > 0.01$ in the Hardy-Weinberg equilibrium test[24]. Of 172 SNPs, 115 met the criteria and the pairwise LD index, Δ (ref. 25), was calculated and plotted (Fig. 1a). We identified six LD blocks with a threshold of $\Delta = 0.5$. We numbered these blocks 1 to 6 from centromere to telomere. Blocks 1–6 spanned roughly 860, 240, 60, 100, 65 and 1 kb, respectively (Fig. 1b and Supplementary Table 1 online).

We divided the SNPs that were in absolute LD relationship with each other ($\Delta > 0.97$) into groups. A total of 90 SNPs was classified into 18 groups (A–R; see Supplementary Table 1 online). We selected representative SNPs by choosing one SNP from each group. We genotyped 18 representative SNPs in 830 case samples and tested association for each by comparison with the genotypes of 658 controls. Two SNPs showed a significant association in recessive trait comparison (Table 1): slc2F1 ($\chi^2 = 17.1$, $P = 0.000034$) and slc5R1 ($\chi^2 = 13.4$, $P = 0.0050$). slc2F1 was located in block 2 and slc5R1 in block 6, and the distance between the two SNPs was over 680 kb. slc2F1 and slc5R1 were in weak LD with $\Delta < 0.15$. When Bonferroni's correction was applied, the P values were 0.00062 and 0.09 for multiple tests of 18 SNPs and 0.0039 and 0.58 for multiple tests of 115 SNPs, respectively. The association with slc2F1 was significant even after the most conservative correction. We therefore concluded that the association in this region was statistically significant and that it originated in block 2.

There were 38 SNPs in block 2, of which 31 were distributed in four known genes (Fig. 1c and Supplementary Table 1 online). Of the 38 SNPs in block 2, 32 belonged to six groups (groups E–J) and the other 6 SNPs did not belong to any group (Fig. 1c and Supplementary Table 1 online). To evaluate further the association with rheumatoid arthritis in block 2, we genotyped the six SNPs that did not belong to any group, in addition to six SNPs representing the six groups E–J, all of which were not associated ($P > 0.05$; data not shown). This analysis implied that, of all of the 38 SNPs in block 2, only the SNPs in group F showed a significant association with rheumatoid arthritis. Group F contained 12 SNPs including slc2F1, 11 of which were clustered in *SLC22A4*, whereas
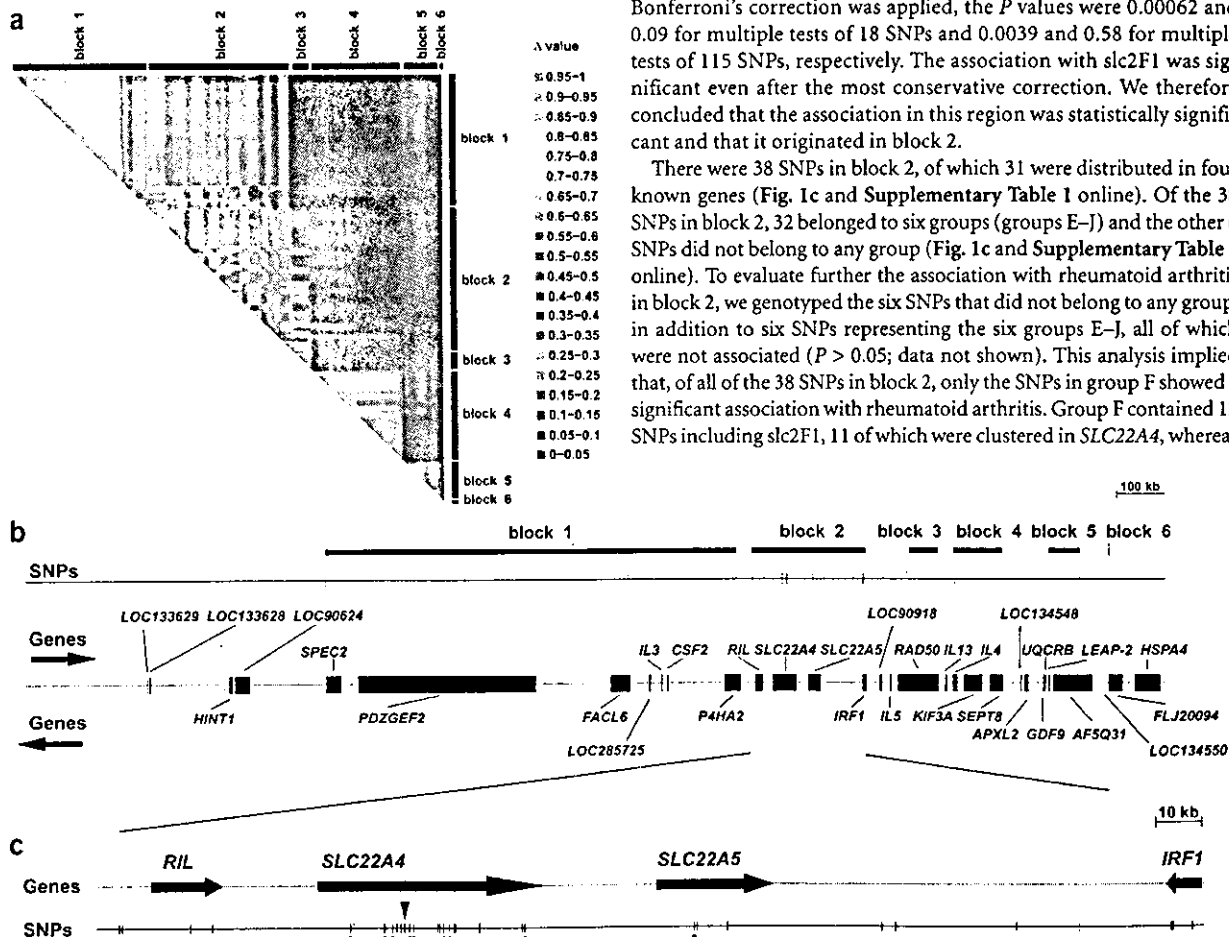
Figure 1 LD blocks and genomic structure of NT_007072.11. (a) Pairwise LD between SNPs, as measured by Δ in 658 controls. The region was divided into six LD blocks. The scale is nominal. (b) Location of genes and 115 SNPs in the NT_007072.11 region. (c) Expanded map of block 2. Arrowhead indicates the location of slc2F1, which shows the strongest association in this region. Dots indicate the SNPs that show strong LD ($\Delta > 0.97$) to slc2F1.
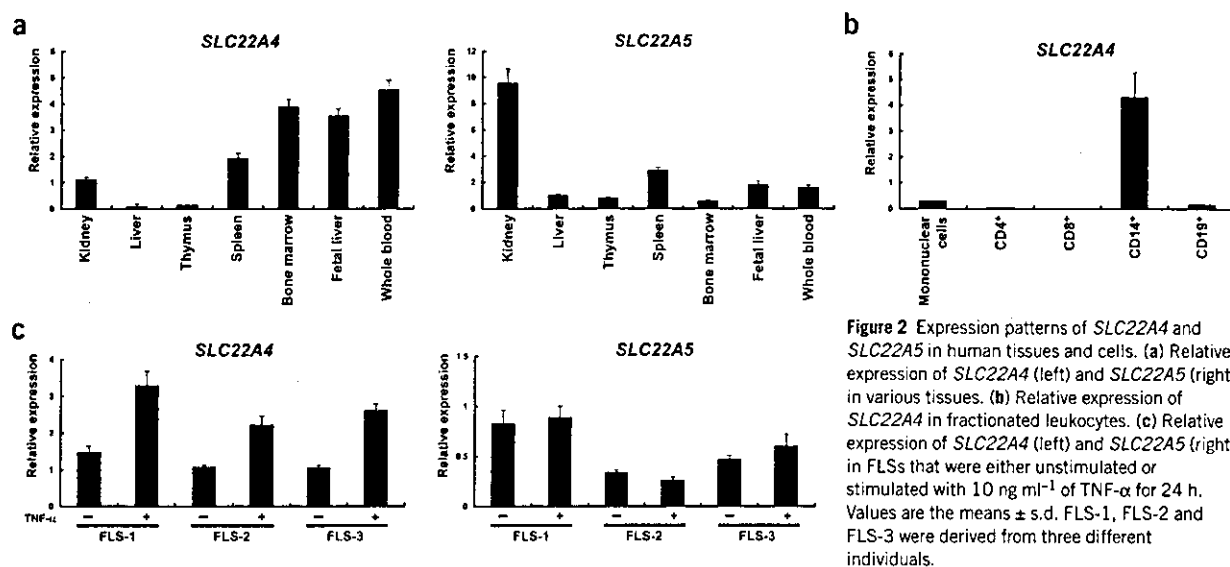
Figure 2 Expression patterns of *SLC22A4* and *SLC22A5* in human tissues and cells. (a) Relative expression of *SLC22A4* (left) and *SLC22A5* (right) in various tissues. (b) Relative expression of *SLC22A4* in fractionated leukocytes. (c) Relative expression of *SLC22A4* (left) and *SLC22A5* (right) in FLSs that were either unstimulated or stimulated with 10 ng ml⁻¹ of TNF-α for 24 h. Values are the means ± s.d. FLS-1, FLS-2 and FLS-3 were derived from three different individuals.

SLC22A5 in block 2 contained only 1 SNP. Neither the gene encoding the LIM domain protein (*RIL*) nor that encoding interferon regulatory factor 1 (*IRF1*) contained any SNP in the same group as slc2F1.

We further analyzed the haplotype constitution of block 2 with the same 12 SNPs. Nine haplotypes were estimated by the expectation maximization algorithm[26] to occur at greater than 2% in both the rheumatoid arthritis and the control groups (Table 2). There are three common haplotypes, denoted 1, 2 and 3. Haplotype 1 was more frequently observed, whereas haplotype 3 was less frequently observed, in the cases than in the controls. The difference in the frequency of haplotype 2 between the case and control groups was marginal. Haplotype 1 was distinguished from the other two common haplotypes by its possession of a susceptibility allele of slc2F1. We therefore assumed that polymorphisms in strong LD with slc2F1 were the origin of association with rheumatoid arthritis in block 2. Association tests using haplotype data were not considered to be use-

ful, because the numbers of haplotypes and diploid types consisting of haplotypes were large.

Expression of *SLC22A4* and *SLC22A5*

We compared the expression patterns of *SLC22A4* and *SLC22A5* using quantitative real-time PCR. *SLC22A4* mRNA was highly expressed in whole blood, bone marrow and fetal liver as compared with kidney (Fig. 2a) and was predominantly expressed in CD14⁺ cells among the peripheral blood mononuclear cells (Fig. 2b); by contrast, *SLC22A5* was almost exclusively expressed in the kidney (Fig. 2a). In fibroblast-like synoviocytes (FLSs) from individuals with rheumatoid arthritis, both *SLC22A4* and *SLC22A5* were moderately expressed (Fig. 2c).

Levels of *SLC22A4* mRNA were increased twofold by stimulation with tumor-necrosis factor-α (TNF-α), but this response was not observed for *SLC22A5* mRNA (Fig. 2c). *SLC22A4* was also expressed in the synovial tissues of individuals with rheumatoid arthritis (Supplementary Fig. 1 online), and the mouse homolog of *SLC22A4* was expressed in inflammatory joints of mice with collagen-induced arthritis, a model of human arthritis, but not in the joints of normal mice (Supplementary Fig. 2 online).
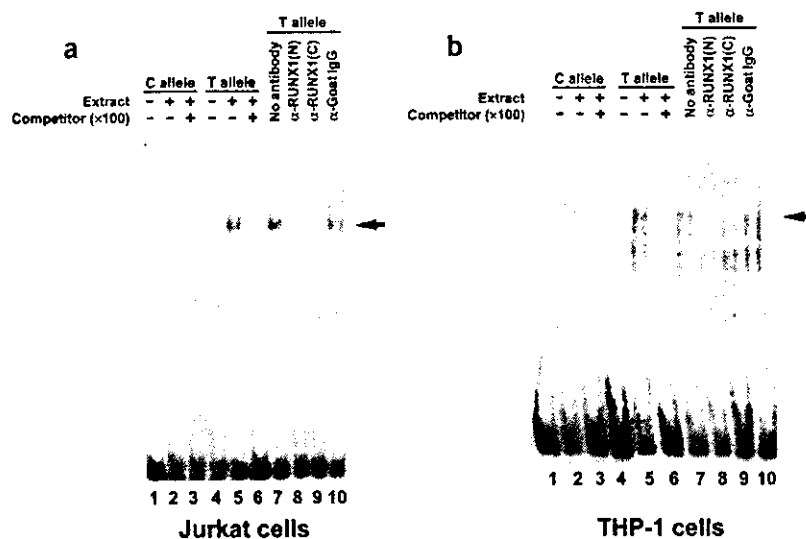
Figure 3 EMSAs and supershift assays with DIG-labeled oligonucleotides of 30 bp corresponding to the slc2F2 polymorphic site. (a) Assays in Jurkat cells; (b) assays in THP-1 cells. For both types of cell, tighter binding of nuclear proteins is observed with the oligonucleotide containing the T allele (lane 5) than with the oligonucleotides containing the C allele (lane 2). In both a and b, antibodies to RUNX1, RUNX1(N) and RUNX1(C) produce a supershifted band (lanes 8 and 9), whereas an antibody to goat IgG does not (lane 10). Binding was competed by a 100-fold excess of unlabeled oligonucleotide (lanes 3 and 6). Arrow indicates the band that was supershifted by the addition of antibodies to RUNX1.

— 280 —

Table 1 Summary of the association of representative SNPs between cases and controls

| SNP ID | Group | Block | Case | | | | | Control | | | | | Allele 1 versus Allele 2 | | | Genotype 11 versus 12+22 | | | Genotype 11+12 versus 22 | | | Δ^a |
			11	12	22	Sum	Minor allele frequency	11	12	22	Sum	Minor allele frequency	Odds ratio (95% c.i.)	χ^2	P	Odds ratio (95% c.i.)	χ^2	P	Odds ratio (95% c.i.)	χ^2	P	
slc1A1	A	1	186	393	243	822	0.47	145	312	192	649	0.46	1.01 (0.87–1.16)	0.01	0.934	1.02 (0.79–1.30)	0.02	0.90	1.01 (0.76–1.35)	0.00	0.99	0.168
slc1B1	B	1	181	395	247	823	0.46	134	319	189	642	0.46	1.01 (0.87–1.17)	0.02	0.883	1.07 (0.83–1.37)	0.27	0.60	1.03 (0.77–1.39)	0.06	0.81	0.180
slc1C1	C	1	148	383	290	821	0.41	103	335	215	653	0.41	1.00 (0.86–1.16)	0.00	0.968	1.17 (0.89–1.55)	1.31	0.25	1.07 (0.78–1.45)	0.93	0.34	0.110
slc1D1	D	1	150	376	298	824	0.41	116	311	221	648	0.42	1.04 (1.20–0.89)^b	0.23	0.631	1.02 (0.78–1.33)	0.02	0.88	1.04 (1.41–0.78)^b	0.67	0.41	0.135
slc2E1	E	2	100	327	392	819	0.32	55	254	332	641	0.28	1.20 (1.02–1.40)	4.85	0.028	1.48 (1.05–2.10)	4.99	0.03	1.54 (1.07–2.21)	2.22	0.14	0.845
slc2G1	G	2	53	321	450	824	0.26	51	261	338	650	0.28	1.11 (1.30–0.94)^b	1.50	0.221	1.24 (1.85–0.83)^b	1.11	0.29	1.28 (1.92–0.85)^b	1.00	0.32	0.482
slc2F1	F	2	132	321	366	819	0.36	58	288	310	656	0.31	1.25 (1.07–1.46)	7.92	0.005	1.98 (1.43–2.75)	17.18	0.000034	1.93 (1.37–2.72)	0.97	0.33	–
slc2H1	H	2	103	372	353	828	0.35	91	282	281	654	0.35	1.02 (1.19–0.88)^b	0.10	0.747	1.14 (1.54–0.84)^b	0.70	0.40	1.11 (1.54–0.81)^b	0.02	0.90	0.482
slc2I1	I	2	70	346	405	821	0.30	77	267	300	644	0.33	1.15 (1.35–0.99)^b	3.22	0.073	1.46 (2.05–1.04)^b	4.70	0.03	1.49 (2.13–1.04)^b	1.09	0.30	0.465
slc2J1	J	2	105	354	361	820	0.34	67	287	300	654	0.32	1.10 (0.95–1.29)	1.59	0.208	1.29 (0.93–1.78)	2.31	0.13	1.30 (0.92–1.83)	0.50	0.48	0.510
slc3K1	K	3	32	282	507	821	0.21	36	205	413	654	0.21	1.01 (1.20–0.84)^b	0.00	0.944	1.44 (2.34–0.88)^b	2.14	0.14	1.39 (2.27–0.84)^b	0.30	0.58	0.097
slc4L1	L	4	82	375	364	821	0.33	74	261	308	643	0.32	1.05 (0.90–1.22)	0.34	0.558	1.17 (1.63–0.84)^b	0.88	0.35	1.06 (1.52–0.75)^b	1.84	0.17	0.149
slc4M1	M	4	50	309	466	825	0.25	35	225	395	655	0.25	1.13 (0.96–1.35)	2.07	0.150	1.14 (0.73–1.78)	0.35	0.56	1.21 (0.77–1.90)	2.19	0.14	0.131
slc4N1	N	4	76	375	344	795	0.33	77	263	307	649	0.32	1.04 (0.75–1.44)	0.16	0.686	1.31 (0.94–1.83)^b	2.55	0.11	1.18 (0.96–1.45)	2.35	0.13	0.147
slc4O1	O	4	106	389	324	819	0.37	88	304	259	651	0.37	1.01 (1.18–0.87)^b	0.01	0.922	1.05 (1.42–0.78)^b	0.10	0.75	1.04 (1.45–0.75)^b	0.01	0.93	0.052
slc5P1	P	5	195	411	213	819	0.49	160	317	175	652	0.49	1.00 (0.87–1.16)	0.00	0.978	1.04 (1.32–0.82)^b	0.11	0.75	1.00 (0.75–1.34)	0.13	0.72	0.088
slc5Q1	Q	5	112	394	305	811	0.38	87	299	258	644	0.37	1.06 (0.91–1.23)	0.58	0.446	1.03 (0.76–1.39)	0.03	0.87	1.09 (0.79–1.51)	0.91	0.34	0.117
slc6R1	R	6	67	356	399	822	0.30	81	268	294	643	0.33	1.18 (1.39–1.01)^b	4.42	0.036	1.62 (2.29–1.15)^b	7.85	0.0051	1.64 (2.33–1.15)^b	1.15	0.28	0.127

^a Δ shows pairwise LD with slc2F1. ^b When the odds ratio is less than 1, an inverted score is indicated.

Because the LD mapping data and expression data of *SLC22A4* and *SLC22A5* suggested that the former gene was more likely to be responsible for the pathological mechanism underlying rheumatoid arthritis, we focused on *SLC22A4* as the susceptibility gene at this locus, although *SLC22A5*, *RIL* and *IRF1* could not be formally ruled out.

An associated SNP affects *SLC22A4* regulation

To identify a causal SNP in the genomic region of *SLC22A4*, we evaluated a sequence extending from 1 kb upstream of the transcription start site to 1 kb downstream of the transcription end site for the presence of SNPs[27]. In addition to the previously registered SNPs, we identified three SNPs in the 5' flanking region (−980A→G, −373C→G and −83G→C) and two missense substitutions in the coding region of *SLC22A4*. None of the three SNPs in the 5' flanking region belonged to group F, one exonic SNP (T306I) was not associated with rheumatoid arthritis (P > 0.05) and the other exonic SNP (G462E) was too rare to be an origin of association of slc2F1 (minor allele frequency < 0.005).

We found nine SNPs in intron 1 that showed a strong LD relation (Δ > 0.97) with slc2F1. For these nine SNPs, we predicted a potential allelic difference in *cis*-acting regulatory function in transcription by reference to the TRANSFAC database[28]. Although four of the nine SNPs were identified to possibly alter transcription factor affinity between the two alleles, the effect of slc2F2 on the affinity of RUNX1 was the most pronounced. The rheumatoid arthritis–susceptible genomic sequence surrounding the slc2F2 T allele (TGTGGT, where the last nucleotide is polymorphic) was predicted to have 100% homology to both a core and a matrix consensus sequence of the RUNX1-binding site, whereas the non-susceptible C allele (TGTGGC) had <80% and <85% homology, respectively.

We examined the allelic difference between the susceptible T allele and the nonsusceptible C allele of slc2F2 in the binding of nuclear proteins from the two hematopoietic cell lines by using electrophoretic mobility shift assay (EMSA). The intensity of the DNA-protein complex from the T allele was higher than that from the C allele in the presence of nuclear extracts from both the Jurkat and THP-1 cells (Fig. 3a,b). Antibodies against RUNX1 supershifted the bands of the DNA-protein complex (Fig. 3a,b). In EMSAs of the nuclear extracts from THP-1 cells, we observed multiple shifted bands, which seemed to be due to the high proteolytic activity of these nuclear extracts[29,30] (Fig. 3b).

To examine the effect of slc2F2 in transcriptional regulation, we transiently transfected cells with either single or five concatenated copies of the sequences surrounding the polymorphic site coupled to a luciferase reporter gene (Fig. 4a). In Jurkat cells, transfection of the susceptible T allele sequence, as compared with the nonsusceptible C allele sequence, resulted in a decrease in luciferase activity of 19% (P < 0.05) in single-copy assays and 38% (P < 0.005) in assays with five concatenated copies. The same trend was observed in THP-1 cells (P < 0.02) in assays with five concatenated copies (Fig. 4b).

Table 2 Haplotype structures of block 2 and their frequencies

	Haplotype frequency		SNP ID (slc2-)												
Haplotype ID	Case	Control	E1	−1	G1	F1	H1	I1	−2	−3	−4	−5	−6	J1
haplotype 1	0.23	0.18	G	C	G	A	C	G	C	T	C	G	A	C
haplotype 2	0.20	0.21	A	T	A	G	C	A	C	C	G	G	A	A
haplotype 3	0.18	0.16	A	T	G	G	G	G	T	C	C	T	C	A
haplotype 4	0.04	0.06	A	T	G	G	G	G	T	C	C	T	A	C
haplotype 5	0.04	0.05	A	T	G	G	G	G	T	C	C	T	A	A
haplotype 6	0.04	0.03	G	C	G	A	C	G	C	T	C	T	C	A

To define further the role of RUNX1 in suppressing transcriptional activity of the sequence surrounding slc2F2, we cotransfected the RUNX1 expression vector with the luciferase constructs containing five concatenated copies of the slc2F2 sequences into HEK293 cells. Overexpression of RUNX1 suppressed the luciferase activity of pT5-slc2F2 and pC5-slc2F2 by 60% and 67%, respectively (see Supplementary Fig. 3 online).

The above data suggest that RUNX1 suppresses the expression of SLC22A4 and that the susceptible T allele tends to express fewer SLC22A4 transcripts owing to a stronger suppressive effect of RUNX1 on this allele.

A SNP in RUNX1 is associated with rheumatoid arthritis

To evaluate whether RUNX1 is also associated with rheumatoid arthritis, we carried out a case-control association test using a SNP of RUNX1. We identified a SNP (denoted runx1) located in intron 6 of RUNX1 that is strongly associated with rheumatoid arthritis (Table 3). Because a functional relation between RUNX1 and SLC22A4 implied that epistasis may be involved, we further evaluated whether there was a specific gene-gene interaction between SNPs in SLC22A4 and RUNX1 (slc2F1 and runx1) by using data from individuals who were genotyped for both SNPs (719 cases and 441 controls).

The genotype that was homozygous with respect to the susceptible alleles of both SNPs showed a high odds ratio (9.03; 95% confidence interval (c.i.) = 2.07–39.38), whereas the genotype that was homozygous with respect to the susceptible allele of SLC22A4 and heterozygous with respect to RUNX1 showed a moderately high odds ratio (2.48; 95% c.i. = 1.37–4.47; Table 4). We investigated the epistatic effect of the two loci by logistic regression[31]. This analysis showed that the loci are independently associated with rheumatoid arthritis with a dominance effect ($P < 0.001$), and a model with epistatic effects did not improve the fit ($P > 0.05$).

DISCUSSION

We examined the cytokine gene cluster in chromosome 5q31 by LD mapping of 115 SNPs and identified a SNP that shows significant association with rheumatoid arthritis and is considered to originate in SLC22A4. This association was identified unexpectedly in SLC22A4 and not in other genes with known immunological functions, but the previous suggestion that SLC22A4 is associated with Crohn disease may support our findings[16]. Because our functional evaluations indicate that SLC22A4 is a susceptibility gene for rheumatoid arthritis, our interpretation of the LD mapping data indicating that SLC22A4 is the origin of the association would seem to be correct.

Most studies on organic cation transporters, including SLC22A4, have been limited to the metabolism and excretion of charged organic substrates in metabolic organs, such as the kidney, liver and placenta.

As an example of the physiological function of nonmetabolic organs, SLC22A3 (also known as OCT3) might be important in the brain as an extraneuronal monoamine transporter[32]. In one study SLC22A4 was suggested to be located in renal epithelial cells, although its precise location in the kidney was not examined[33], whereas in a conflicting report the expression of SLC22A4 mRNA was found to be extremely low in the kidney[34]. We showed that SLC22A4 is expressed in immunological and hematological organs and tissues, and also that SLC22A4 mRNA is induced by proinflammatory stimuli. These results suggest that SLC22A4 functions as a transporter in lymphoid organs or inflammatory milieu.
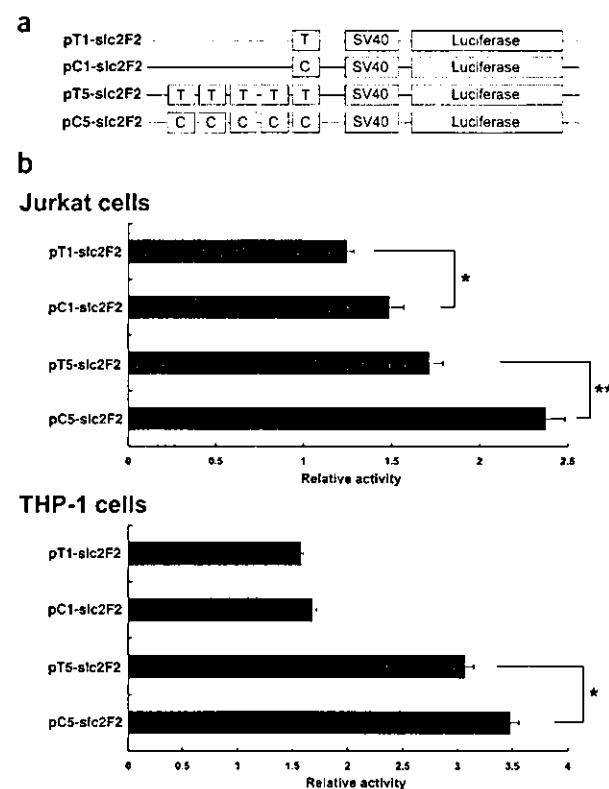

Figure 4 Comparison of allelic variants of slc2F2 analyzed by relative luciferase activity. (a) Luciferase constructs. (b) Allelic differences in relative luciferase activity in Jurkat cells (top) and THP-1 cells (bottom). Relative activity was calculated by taking the relative luciferase activity of the empty vector (pGL3-promoter) to be 1. Data show the mean ± s.e.m. relative activity from four experiments done in quadruplicate. *$P < 0.05$ and **$P < 0.005$ by Student's t-test.

Table 3 Association of runx1 between cases and controls

SNP ID	Gene structure	Location*	Genotype Case CC	CG	GG	Sum	Minor allele frequency	Control CC	CG	GG	Sum	Minor allele frequency	C allele versus G allele Odds ratio (95% c.i.)	χ^2	P	Genotype CC versus CG+GG Odds ratio (95% c.i.)	χ^2	P	Genotype CC+CG versus GG Odds ratio (95% c.i.)	χ^2	P
runx1	intron 6	24658	140	423	257	820	0.43	93	295	262	650	0.37	1.28 (1.10–1.48)	10.37	0.0013	1.23 (0.93–1.64)	2.08	0.15	1.48 (1.19–1.83)	12.76	0.00035

*Location represents nucleotide positions counted from the nucleotide adjacent to the 3' end of exon 6.

To determine the SNPs responsible for susceptibility to rheumatoid arthritis, we considered that the location of SNPs in the gene structure of *SLC22A4* would be important for identifying functional variants that cause susceptibility. None of the SNPs in the coding or 5' flanking region of *SLC22A4* were associated with rheumatoid arthritis. As several studies have reported that transcriptional regulatory elements are located in introns[12,35–37], we evaluated SNPs in the first intron of *SLC22A4* for their roles in transcriptional regulation and found that a genomic DNA sequence containing the slc2F2 T→C polymorphism is located in a RUNX1 consensus sequence (TGTGGT).

The RUNX1 transcription factor is expressed mainly in hematopoietic cells and functions both to activate and to repress transcription through interactions with cofactors[20]; it is also well known to be a primary factor in acute myeloid leukemia[21]. Several promoters of hematopoietic genes, such as *IL3* and *CSF2*, have been suggested to be regulated by RUNX1 (refs. 38,39). Notably, an intronic SNP of programmed cell death gene 1 (*PDCD1*) is reported to cause susceptibility to the autoimmune disease systemic lupus erythematosus, and this SNP alters a binding site for RUNX1. But the effect of RUNX1 on *PDCD1* expression has yet to be defined[22]. This example is very similar to our findings in rheumatoid arthritis: a gene (*SLC22A4*) with a susceptibility variant for this disorder and a regulatory molecule (RUNX1) with an allelic difference for this susceptibility variant.

In a preliminary test, we examined the expression of *SLC22A4* mRNA in whole blood from healthy volunteers versus each genotype of slc2F1, but the results were not conclusive owing to large interindividual variation and a small sample size (data not shown). We show here, however, that the suppressive effect of RUNX1 on *SLC22A4* expression is affected by the intronic SNP slc2F2, and that this suppression seems to be stronger with the susceptible T allele than with the nonsusceptible C allele, leading to a lower expression of *SLC22A4*. Although a functional difference between the two alleles of the disease-associated SNP in *RUNX1* has not yet been confirmed, there seems to be an association between the SNP in *RUNX1* and rheumatoid arthritis.

We propose the following scheme: first, RUNX1 binds as an inhibitory factor to the DNA sequence containing slc2F2; second, although *SLC22A4* expression is enhanced in inflammatory conditions, it is negatively regulated by RUNX1; and third, an excessive suppression of *SLC22A4* causes susceptibility to rheumatoid arthritis. Although inappropriately suppressed *SLC22A4* expression seems to be associated with rheumatoid arthritis, the mechanisms underlying where and how this *SLC22A4* insufficiency contributes to the pathophysiology of rheumatoid arthritis are unclear.

As *SLC22A4* is expressed in both lymphoid organs and arthritic joints, there seem to be two reasonable hypotheses. First, *SLC22A4* may be related to rheumatoid arthritis in the development of autoimmunity in secondary lymphoid organs. Second, *SLC22A4* may contribute to inflammatory processes in inflammatory joints. The fact that transcription of *SLC22A4* is regulated by RUNX1 may support the first hypothesis because the primary function of RUNX1 is known to be related to the differentiation of hematopoietic cells[20]. The increased expression of *SLC22A4* in inflamed joints, however, supports the second hypothesis.

To investigate further the role of *SLC22A4* in rheumatoid arthritis pathogenesis, a report on neutrophil cytosolic factor 1 (*Ncf1*) in rat arthritis may be useful. *Ncf1* has been identified as a gene that regulates the severity of arthritis in rats[40]. Because *Ncf1* regulates the oxidative burst in neutrophils, it seems reasonable to speculate that it may be related to rheumatoid arthritis by affecting inflammatory processes. It turns out, however, that the disease-related allele of *Ncf1* promotes the activation of arthritogenic T cells in lymphoid organs, rather than in peripheral inflammatory joints, through an oxygen burst response. Further analysis is required to specify the substrates of SLC22A4 to establish its physiological role in inflammation and in the pathogenesis of rheumatoid arthritis.

In summary, we identified *SLC22A4* as a susceptibility gene for rheumatoid arthritis through a case-control association study with SNPs. We characterized the expression of *SLC22A4* in hematopoietic cells, as well as the functional SNP that has an allele-specific effect on expression through varying affinity for the transcription factor RUNX1. In addition, a SNP in the *RUNX1* is also significantly associated with rheumatoid arthritis. These findings suggest that the polymorphisms of

Table 4 Case-control genotypes for slc2F1 and runx1

		Genotype of runx1 Cases (n = 719) CC	GC	GG	Controls (n = 441) CC	GC	GG	slc2F1/runx1 Odds ratio (95% c.i.) CC	GC	GG
Genotype of slc2F1	AA	23	60	33	2	19	17	9.03 (2.07–39.38)	2.48 (1.37–4.47)	1.52 (0.79–2.92)
	AG	49	148	83	28	82	85	1.37 (0.80–2.37)	1.42 (0.96–2.10)	0.77 (0.51–1.16)
	GG	48	168	107	26	98	84	1.45 (0.83–2.53)	1.35 (0.92–1.97)	1

both genes affect *SLC22A4* gene expression and thus susceptibility to rheumatoid arthritis. This might be an example of a complex genetic trait created by multiple susceptibility genes.

METHODS

Subjects. A total of 830 individuals with rheumatoid arthritis (84.2% women; mean age ± s.d., 59.0 ± 12.3 years; 75.0% rheumatoid factor positive) and 658 controls (52% women; mean age ± s.d., 49.5 ± 22.1 years) was recruited through several medical institutes in Japan. We sampled synovial tissues from individuals with rheumatoid arthritis who underwent arthroplastic surgery. All individuals satisfied the American Rheumatism Association's revised criteria for classification of rheumatoid arthritis[41]. All subjects were Japanese and consented to participate in the study in accordance with the process approved by the Ethical Committee at the SNP Research Center, The Institutes of Physical and Chemical Research (RIKEN), Yokohama, Japan. We extracted genomic DNA from peripheral blood leukocytes of affected individuals and controls subjects with standard protocols[12]. We extracted total RNA from whole blood by the PAXgene blood RNA system (Qiagen).

SNPs and genotyping methods. We extracted 172 SNPs located in NT_007072.11 (NCBI) from the JSNP database[23]. SNPs in *SLC22A4* were identified by directly sequencing DNA from 48 Japanese individuals[27]. We genotyped a SNP, rs2268277 (NCBI), in *RUNX1*. SNPs were genotyped by using the Invader assay[11] and the TaqMan assay[13]. The probe sets for the Invader assay were designed and synthesized by Third Wave Technologies and those for the TaqMan assay were obtained from Applied Biosystems.

Quantitative real-time PCR. Preparations of rheumatoid FLSs have been described[42]. We treated FLSs between passages 10 and 15 with 10 ng ml⁻¹ of TNF-α (Genzyme) for 24 h (or left them untreated) and isolated total cellular RNA by an RNeasy kit (Qiagen). We purchased the Human Blood Fractions MTC Panel and human total RNA from various tissues from Clontech. TaqMan probes and primers for *SLC22A4*, *SLC22A5* and *GAPDH* were Assay-on-Demand gene expression products (Applied Biosystems). We synthesized first-strand cDNA with 1 μg of total RNA using oligo $d(T)_{16}$ primers and TaqMan Reverse Transcription Reagents (Applied Biosystems).

TaqMan PCR was done with an ABI PRISM 7000 Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions. The relative expression of *SLC22A4* and *SLC22A5* mRNA was normalized to the amount of *GAPD* in the same cDNA by using the standard curve method described by the manufacturer.

EMSA. We prepared nuclear extracts from Jurkat cells and THP-1 cells (RIKEN cell bank) using NE-PER extraction reagents (Pierce). We stored nuclear extracts in aliquots at –80 °C until use. Protein concentration was measured by the BCA Protein Assay kit (Pierce) with bovine serum albumin as a standard. EMSAs were done with the DIG Gel Shift kit (Roche Diagnostics) using digoxigenin (DIG)-labeled double-stranded oligonucleotides specific for the slc2F2 C allele and the slc2F2 T allele. We incubated DIG-labeled probes with nuclear extracts for 30 min and separated them by electrophoresis on a 5% nondenaturing polyacrylamide gel with 0.5-fold TBE running buffer. DNA-protein complexes were electroblotted to nylon membrane and the bands shift were visualized according to the user's manual for the DIG Gel Shift kit. For the supershift assay, we incubated protein-DNA complexes with 1 μg of antibodies specific for the N terminus, RUNX1(N), or the C terminus, RUNX1(C), of RUNX1 (AML-1 (N-20) and AML-1 (C-19), respectively; Santa Cruz Biotech) for 30 min at 4 °C before electrophoresis. Oligonucleotide sequences for EMSA are available from the authors on request.

Luciferase assay. We constructed luciferase reporter plasmids by cloning single or five concatenated copies of the adjacent 24 nucleotides of slc2F2 into the pGL3-promoter vector (Promega) upstream of the SV40 promoter. For experimental convenience, we introduced *NheI* or *XbaI* sites at the 5' end of sense or antisense oligonucleotides, respectively, and added thymine or guanine to the 3' end of sense or antisense oligonucleotides, respectively. We sequenced the inserted portions of all constructs to verify the nucleic acid sequences and the identity and location of the SNP.

For the luciferase assay, we cultured Jurkat cells and THP-1 cells in RPMI1640 medium supplemented with 10% fetal bovine serum and maintained HEK293 cells (obtained from the RIKEN cell bank) in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Jurkat cells were transfected with 0.8 μg of either one of the constructs and 0.2 μg of the pRL-TK *Renilla* luciferase vector (Promega) as an internal control for transfection efficiency by using 2 μl of Lipofectamine 2000 reagent (Invitrogen). After 6 h, we stimulated the transfected cells with phorbol myristate acetate (20 ng ml⁻¹) and ionomycin (1 μM). THP-1 cells were transfected with 0.49 μg of either one of the constructs and 0.01 μg of pRL-TK by using Effectene reagent (Qiagen). After 24 h, we collected the cells and measured luciferase activity with the Dual-Luciferase Reporter Assay System (Promega).

Mathematical and statistical analyses. We assessed association and Hardy-Weinberg equilibrium by the χ^2 test and Fisher's exact test with Bonferroni's correction. Haplotype frequencies were estimated using the expectation maximization algorithm[26]. The LD index (Δ) was calculated in 658 unaffected individuals[25], and Figure 1a was drawn by an application created by our group in Excel (Microsoft). Luciferase assay data were assessed by Student's *t*-test using Excel (Microsoft).

Note: Supplementary information is available on the Nature Genetics website.

1. Seldin, M.F., Amos, C.I., Ward, R. & Gregersen, P.K. The genetics revolution and the assault on rheumatoid arthritis. *Arthritis Rheum.* **42**, 1071–1079 (1999).
2. Deighton, C.M., Walker, D.J., Griffiths, I.D. & Roberts, D.F. The contribution of HLA to rheumatoid arthritis. *Clin. Genet.* **36**, 178–182 (1989).
3. Nepom, G.T. Major histocompatibility complex-directed susceptibility to rheumatoid arthritis. *Adv. Immunol.* **68**, 315–332 (1998).
4. Cornelis, F. *et al.* New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc. Natl. Acad. Sci. USA* **95**, 10746–10750 (1998).
5. Jawaheer, D. *et al.* A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am. J. Hum. Genet.* **68**, 927–936 (2001).
6. MacKay, K. *et al.* Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum.* **46**, 632–639 (2002).
7. Jawaheer, D. *et al.* Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum.* **48**, 906–916 (2003).
8. Shiozawa, S. *et al.* Identification of the gene loci that predispose to rheumatoid arthritis. *Int. Immunol.* **10**, 1891–1895 (1998).
9. Tabor, H.K., Risch, N.J. & Myers, R.M. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* **3**, 391–397 (2002).
10. Ardlie, K.G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **3**, 299–309 (2002).
11. Ohnishi, Y. *et al.* A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46**, 471–477 (2001).
12. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
13. Suzuki, A. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
14. Goldbach-Mansky, R. *et al.* Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. *Arthritis Res.* **2**, 236–243 (2000).
15. Grunig, G. *et al.* Requirement for IL-13 independently of IL-4 in experimental asthma. *Science* **282**, 2261–2263 (1998).
16. Rioux, J.D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29**, 223–228 (2001).

— 284 —

ARTICLES

17. Mansur, A.H., Bishop, D.T., Markham, A.F., Britton, J. & Morrison, J.F. Association study of asthma and atopy traits and chromosome 5q cytokine cluster markers. *Clin. Exp. Allergy* **28**, 141–150 (1998).
18. Kauppi, P. *et al.* A second-generation association study of the 5q31 cytokine gene cluster and the interleukin-4 receptor in asthma. *Genomics* **77**, 35–42 (2001).
19. Lee, J.K., Park, C., Kimm, K. & Rutherford, M.S. Genome-wide multilocus analysis for immune-mediated complex diseases. *Biochem. Biophys. Res. Commun.* **295**, 771–773 (2002).
20. Lutterbach, B. & Hiebert, S.W. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene* **245**, 223–235 (2000).
21. Miyoshi, H. *et al.* t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, *AML1. Proc. Natl. Acad. Sci. USA* **88**, 10431–10434 (1991).
22. Prokunina, L. *et al.* A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat. Genet.* **32**, 666–669 (2002).
23. Haga, H., Yamada, R., Ohnishi, Y., Nakamura, Y. & Tanaka, T. Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. *J. Hum. Genet.* **47**, 605–610 (2002).
24. Nielsen, D.M., Ehm, M.G. & Weir, B.S. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.* **63**, 1531–1540 (1998).
25. Devlin, B. & Risch, N. A comparison of linkage disequilibrium measures for finescale mapping. *Genomics* **29**, 311–322 (1995).
26. Ott, J. Counting methods (EM algorithm) in human pedigree analysis: linkage and segregation analysis. *Ann. Hum. Genet.* **40**, 443–454 (1977).
27. Saito, S. *et al.* Catalog of 238 variations among six human genes encoding solute carriers (hSLCs) in the Japanese population. *J. Hum. Genet.* **47**, 576–584 (2002).
28. Heinemeyer, T. *et al.* Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.* **26**, 362–367 (1998).
29. Galson, D.L. & Housman, D.E. Detection of two tissue-specific DNA-binding proteins with affinity for sites in the mouse β-globin intervening sequence 2. *Mol. Cell. Biol.* **8**, 381–392 (1988).
30. Pahl, H.L. *et al.* The proto-oncogene *PU.1* regulates expression of the myeloid-specific CD11b promoter. *J. Biol. Chem.* **268**, 5014–5020 (1993).
31. Cordell, H.J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet* **11**, 2463–2468 (2002).
32. Wu, X. *et al.* Identity of the organic cation transporter OCT3 as the extraneuronal monoamine transporter (uptake2) and evidence for the expression of the transporter in the brain. *J. Biol. Chem.* **273**, 32776–32786 (1998).
33. Yabuuchi, H. *et al.* Novel membrane transporter OCTN1 mediates multispecific, bidirectional, and pH-dependent transport of organic cations. *J. Pharmacol. Exp. Ther.* **289**, 768–773 (1999).
34. Motohashi, H. *et al.* Gene expression levels and immunolocalization of organic ion transporters in the human kidney. *J. Am. Soc. Nephrol.* **13**, 866–874 (2002).
35. Surinya, K.H., Cox, T.C. & May, B.K. Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5-aminolevulinate synthase 2 gene. *J. Biol. Chem.* **273**, 16798–16809 (1998).
36. Ghayor, C. *et al.* Regulation of human *COL2A1* gene expression in chondrocytes. Identification of C-Krox-responsive elements and modulation by phenotype alteration. *J. Biol. Chem.* **275**, 27421–27438 (2000).
37. Beohar, N. & Kawamoto, S. Transcriptional regulation of the human nonmuscle myosin II heavy chain-A gene. Identification of three clustered *cis*-elements in intron-1 which modulate transcription in a cell type- and differentiation state-dependent manner. *J. Biol. Chem.* **273**, 9168–9178 (1998).
38. Uchida, H., Zhang, J. & Nimer, S.D. AML1A and AML1B can transactivate the human IL-3 promoter. *J. Immunol.* **158**, 2251–2258 (1997).
39. Takahashi, A. *et al.* Positive and negative regulation of granulocyte-macrophage colony-stimulating factor promoter activity by AML1-related transcription factor, PEBP2. *Blood* **86**, 607–616 (1995).
40. Olofsson, P. *et al.* Positional identification of *Ncf1* as a gene that regulates arthritis severity in rats. *Nat. Genet.* **33**, 25–32 (2003).
41. Arnett, F.C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
42. Miyazawa, K., Mori, A., Yamamoto, K. & Okudaira, H. Transcriptional roles of CCAAT/enhancer binding protein-β, nuclear factor-κB, and C-promoter binding factor 1 in interleukin (IL)-1β-induced IL-6 synthesis by human rheumatoid fibroblast-like synoviocytes. *J. Biol. Chem.* **273**, 7620–7627 (1998).

nature
genetics

Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis

Akari Suzuki[1], Ryo Yamada[1], Xiaotian Chang[1], Shinya Tokuhiro[1,2], Tetsuji Sawada[3], Masakatsu Suzuki[2], Miyuki Nagasaki[2], Makiko Nakayama-Hamada[2], Reimi Kawaida[2], Mitsuru Ono[2], Masahiko Ohtsuki[2], Hidehiko Furukawa[2], Shinichi Yoshino[4], Masao Yukioka[5], Shigeto Tohma[6], Tsukasa Matsubara[7], Shigeyuki Wakitani[8], Ryota Teshima[9], Yuichi Nishioka[10], Akihiro Sekine[11], Aritoshi Iida[11], Atsushi Takahashi[12], Tatsuhiko Tsunoda[12], Yusuke Nakamura[11,13] & Kazuhiko Yamamoto[1,3]

Individuals with rheumatoid arthritis frequently have autoantibodies to citrullinated peptides, suggesting the involvement of the peptidylarginine deiminases citrullinating enzymes (encoded by *PADI* genes) in rheumatoid arthritis. Previous linkage studies have shown that a susceptibility locus for rheumatoid arthritis includes four *PADI* genes but did not establish which *PADI* gene confers susceptibility to rheumatoid arthritis. We used a case-control linkage disequilibrium study to show that *PADI* type 4 is a susceptibility locus for rheumatoid arthritis ($P = 0.000008$). *PADI4* was expressed in hematological and rheumatoid arthritis synovial tissues. We also identified a haplotype of *PADI4* associated with susceptibility to rheumatoid arthritis that affected stability of transcripts and was associated with levels of antibody to citrullinated peptide in sera from individuals with rheumatoid arthritis. Our results imply that the *PADI4* haplotype associated with susceptibility to rheumatoid arthritis increases production of citrullinated peptides acting as autoantigens, resulting in heightened risk of developing the disease.

Rheumatoid arthritis is one of the most common human systemic autoimmune diseases. It is characterized by inflammation of synovial tissues and the formation of rheumatoid pannus, which is capable of eroding adjacent cartilage and bone and causing subsequent joint destruction. Previous studies have indicated that risk of the disease in siblings of affected individuals (λ_{sib}) is 2–17 times higher, suggesting the importance of genetic factors in rheumatoid arthritis[1]. Multiple genes are believed to contribute to rheumatoid arthritis susceptibility, but the only locus that has been conclusively associated with the condition is the *HLA-DRB* locus, which accounts for about one third of the genetic component[2–4]. Recently, four sibling-pair linkage studies from Europe, North America and Japan were published[5–8]. Although no common loci apart from the *HLA* region were suggested by all the studies, some were suggested by multiple studies. Chromosome 1p36 represents one such locus. Cornelis *et al.*[5] reported an association between rheumatoid arthritis and *D1S228* that identified nucleotides 363,575–363,702 on NT_004873.12 in a study using 114 sibling pairs ($P = 0.0065$). Shiozawa *et al.*[8] obtained a single-point lod score of 3.58 at *D1S214* that identified

nucleotides 1,089,077–108,972 on NT_028054.9 and also observed lod scores of 3.77 as a single-point analysis and 6.13 as a multi-point analysis at *D1S253* that identified a region 1.5 cM telomeric from *D1S214* (located in GB4 map by the International RH Mapping Consortium but not annotated in the Reference Sequence of genomic DNA by NCBI), using 41 families. *D1S228* and *D1S214* are located 6.7 Mb apart according to the Reference Sequence build 33 from the National Center for Biotechnology Information.

The gene region located 3.1 Mb and 9.8 Mb centromeric from *D1S228* and *D1S214*, respectively, contains clusters of enzymes that are functionally associated with the production of rheumatoid arthritis–specific autoantibodies. These enzymes are the peptidylarginine deiminases (PADIs), which posttranslationally convert arginine residues to citrulline. Citrullinated epitopes involved in a peptidic link are the most specific targets of rheumatoid arthritis–specific autoantibodies. Citrullination is related to two rheumatoid arthritis–specific autoantibody systems: those directed against perinuclear factor/keratin and against Sa[9,10]. Assays of antibodies to citrullinated peptide can

be used as valuable diagnostic tools[11,12]. The clinical importance of measuring antibodies to citrullinated peptide and the specificity of autoantibodies suggests a specific role of citrullination and PADIs in the pathophysiology of rheumatoid arthritis. In addition, the appearance of antibodies to citrullinated peptide in sera from affected individuals in the very early phase of disease manifestation implies that citrullination is involved in the triggering phase or the acute phase of the disease[13]. The presence of citrullinated peptides in rheumatoid arthritis synovial tissue has also been reported, suggesting the involvement of PADIs in the pathomechanisms of rheumatoid arthritis[14–16].

We carried out a case-control association study using single-nucleotide polymorphisms (SNPs) discovered by the Japanese Millennium Genome Project in the 1p36 region containing the genes *PADI1*, *PADI2*, *PADI3* and *PADI4*. This study identified a haplotype associated with susceptibility to rheumatoid arthritis in *PADI4* but not in neighboring *PADI* genes. We confirmed that *PADI4* was expressed in hematological cells by northern-blot hybridization and in synovial tissue of individuals with rheumatoid arthritis by *in situ* RT–PCR and immunohistochemistry. Moreover, the susceptibility haplotype of *PADI4* was related to levels of antibody to citrullinated filaggrin in sera of individuals with rheumatoid arthritis. We also identified a difference in mRNA stability between non-susceptibility and susceptibility variants of *PADI4*.

RESULTS

Case-control study using SNPs in NT_034376.1

To identify genes associated with susceptibility to rheumatoid arthritis, we focused on the region NT_034376.1 on chromosome 1p36, in which we had previously identified the SNP strongly associated with rheumatoid arthritis. This region contains eight genes (including four *PADI* genes) that could be associated with rheumatoid arthritis according to the data regarding antibodies to citrullinated peptides. We refined the location of the rheumatoid arthritis susceptibility locus in a case-control study using 119 SNPs distributed in genes across contig NT_034376.1 (Fig. 1a,b and Supplementary Table 1 online). The total length we evaluated was 445,670 bp, and SNPs were located every 3.7 kb on average. We predominantly used the Invader assay, which can efficiently detect genotypes of SNPs[17,18], and analyzed samples from a total of 830 affected individuals and 736 unaffected controls. Overall success rates of typing assays for cases and controls were 96% and 95%, respectively. A SNP in *PADI4*, padi4_94 (28017T in intron 3, susceptible; →C, non-susceptible), had the most significant association with rheumatoid arthritis (χ^2 = 19.856, P = 0.000008 comparing allele 1 versus allele 2; odds ratio (OR) = 1.97, 95% confidence interval (c.i.) = 1.44–2.69 comparing susceptible homozygotes versus non-susceptible homozygotes; Table 1 and Fig. 1b). When Bonferroni's correction was applied to the result we obtained P = 0.00095, and the Monte Carlo Permutation test gave P = 0.00003 with 1×10^6 replications[19]. Both of these results were statistically significant.

We then sequenced all exons of *PADI4*, including the 5′ and 3′ untranslated regions, from 48 individuals with rheumatoid arthritis to identify SNPs. We identified four new SNPs and genotyped them in the exons: padi4_89 (163G→A in exon 2), padi4_90 (245T→C in exon 2), padi4_92 (335G→C in
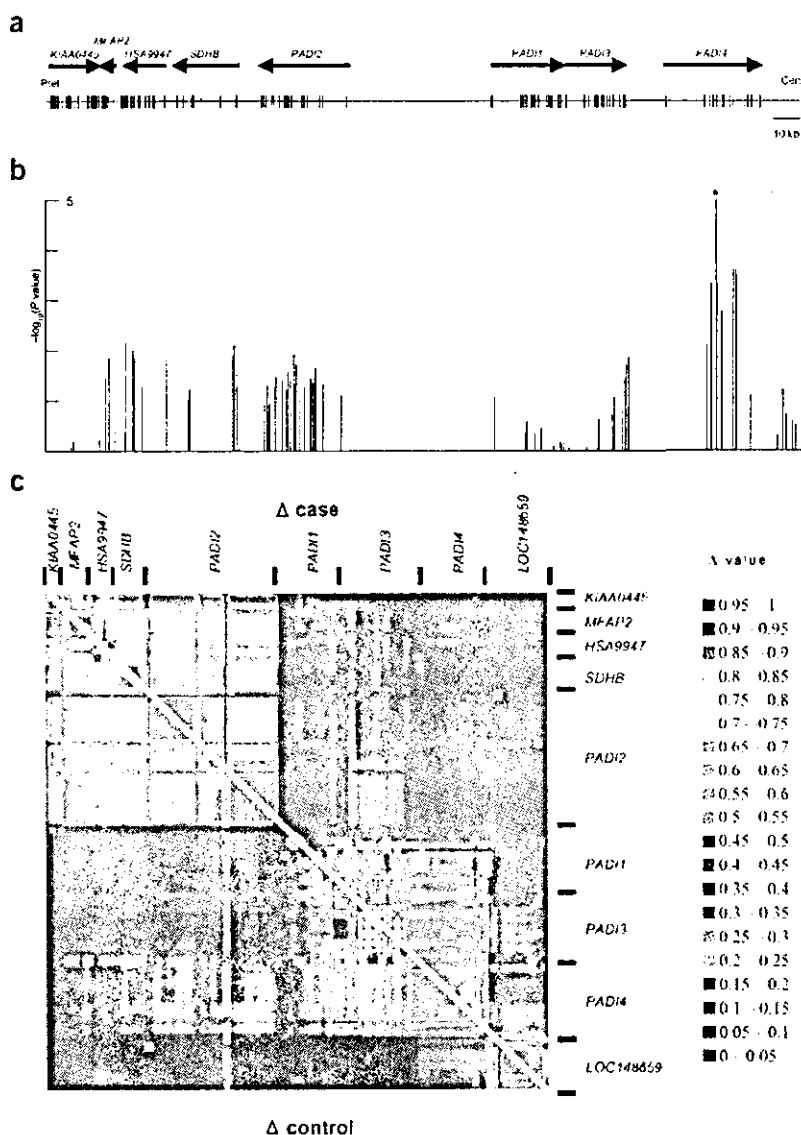
Figure 1 Gene content of NT_034367.1 in chromosome 1p36, case-control association and linkage disequilibrium. (a) Genomic structure of genes in this region. Ptel, p telomere; Cen, centromere. (b) Case-control association plots (−log₁₀(P value)) versus location in this region. Asterisk indicates the SNP showing the strongest association in this region. (c) Pairwise linkage disequilibrium between SNPs, as measured by Δ in the case and control populations in this region: upper right triangle, case population; lower left triangle, control population.

— 287 —

Table 1 Summary of association between cases and controls in *PADI4*

SNP ID	Genotype of case				Genotype of control				Allele 1 versus allele 2		Genotype 11 versus genotype 22
	11	12	22	Sum	11	12	22	Sum	χ^2	P value	OR (95% c.i.)
padi4_92	166	416	241	823	102	307	246	655	12.36	0.00046	1.66 (1.23–2.25)
padi4_94	167	415	240	822	89	305	252	646	19.86	0.0000084	1.97 (1.44–2.69)
padi4_104[a]	268	355	110	733	313	358	64	735	12.67	0.00051	2.00 (1.41–2.86)[b]
padi4_95	131	386	304	821	64	300	281	645	12.29	0.00046	1.89 (1.35–2.66)
padi4_97	304	390	131	825	283	305	64	652	12.48	0.00041	1.92(1.35–2.70)[b]
padi4_99	225	421	181	827	224	331	100	655	13.72	0.00021	1.82 (1.33–2.44)[b]
padi4_100	225	418	180	823	216	332	98	646	12.00	0.00053	1.75 (1.30–2.38)[b]
padi4_101	222	417	178	817	216	322	95	633	13.62	0.00022	1.82 (1.33–2.50)[b]

Sum of cases > 800; *P* < 0.001.
[a]Control sample number of this SNP was 736. [b]For OR >1, the inverted score is indicated.

exon 3) and padi4_104 (349T→C in exon 4; **Table 1** and **Fig. 2a,b**). Overall, eight SNPs in NT_034376.1 had significant associations with rheumatoid arthritis (*P* < 0.001, **Table 1**), and all these SNPs were in *PADI4*. In the case and control populations, strong linkage disequilibrium extended only within *PADI4* and not to SNPs flanking *PADI4* (**Fig. 1c**). We therefore concluded that the strong association detected with SNPs in *PADI4* originated from *PADI4* itself. Rheumatoid factor status, sex, age at disease onset and *HLA-DRB1* status of affected individuals were not related to *PADI4* genotype distribution (data not shown).

We next undertook full haplotype analysis for 17 SNPs in *PADI4*. Only 4 of 2^{17} possible haplotypes were estimated to have frequency >0.02 in both case and control groups using the expectation-maximization algorithm. Less frequently occurring haplotypes were not shown, owing to concern over the accuracy of low frequency alleles in the expectation-maximization algorithm. The most frequently occurring haplotype, haplotype 1, and the second most frequently occurring haplotype, haplotype 2, comprised more than 85% of total chromosomes both in case and control groups (**Table 2**). Among the SNPs that segregate haplotype 1 and haplotype 2, four were exonic and three of them involved amino acid substitutions: padi4_89, padi4_90, padi4_92 and padi4_104, resulting in G55S, V82A, G112A and L117L, respectively (**Fig. 2c**). Haplotype 1 was more frequently observed in the control group and haplotype 2 in the case group. Haplotype 1 and its transcript and peptide were therefore termed 'non-susceptible', and haplotype 2 and its transcript and peptide 'susceptible'. Compositions of bases and amino acids of transcripts and peptides for susceptible and non-susceptible types are indicated in **Figure 2c**.

Expression of *PADI4* mRNA

To investigate the expression patterns of *PADI4* in tissues, we carried out northern-blot analysis and quantitative real-time RT–PCR. Northern-blot analysis identified two *PADI4* transcripts, one band at 2.6 kb and the other at 4.0 kb (**Fig. 3a**), as described in a previous study[20]. *PADI4* had high levels of expression in bone marrow

and peripheral blood leukocytes, low levels of expression in spleen and fetal liver and no expression in other organs (including liver and kidney). *PADI4* was thus highly expressed in the organs of the hematological system.

We also confirmed *PADI4* expression in hematological cell types. Quantitative RT–PCR was done using RNA from CD4$^+$ and CD8$^+$ T cells, CD19$^+$ B cells, CD14$^+$ monocytes, polymorphonuclear leukocytes (PMNs), bone marrow and kidney (as a negative control). *PADI4* was highly expressed in bone marrow, CD14$^+$ monocytes and PMNs but was not expressed in CD4$^+$ and CD8$^+$ T cells or CD19$^+$ B cells (**Fig. 3b**).

Localization of *PADI4* mRNA, protein and citrullinated peptide

To test whether *PADI4* was expressed in rheumatoid arthritis synovial tissues, we carried out *in situ* RT–PCR. We observed *PADI4* mRNA in the lining or sublining layers of synovial tissues from all seven individuals with rheumatoid arthritis that we tested (**Fig. 3c**).
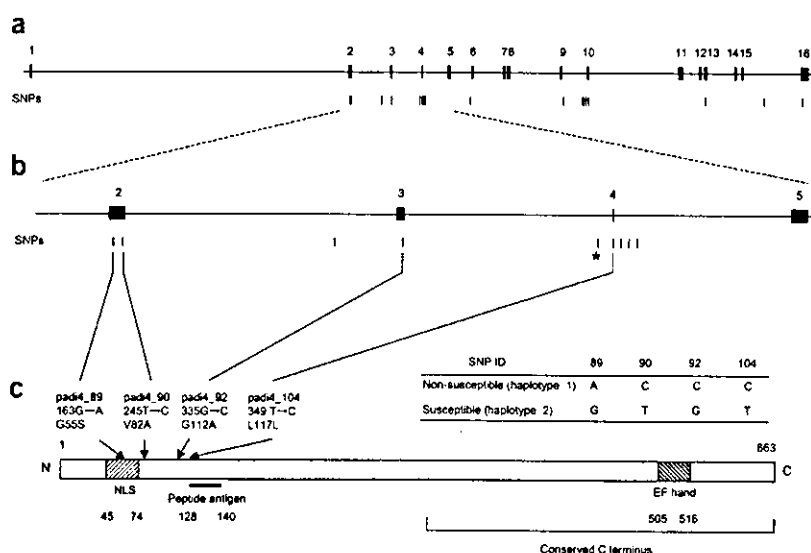

Figure 2 Structure of *PADI4*. (a) Exon-intron structure of *PADI4*. SNPs in *PADI4* are indicated below the gene. (b) Structure of region including exons 2–5. SNPs in this region are indicated below the gene. The asterisk marks the same SNP that is indicated in **Figure 1b**. (c) Protein structure of PADI4. Nucleotide numbering starts from start codons of genes. The bracketed region was used to generate the peptide antibody used in immunohistochemistry.

Table 2 Haplotype structure and frequency in PADI4

| Haplotype ID | Haplotype frequency | | SNP ID (as padi4_x) | | | | | | | | | | | | | | | | | |
	Case	Control	89	90	91	92	93	94	104	95	96	97	98	99	100	101	102	103	105
Haplotype 1	0.52	0.60	A	C	C	C	C	C	C	G	T	T	C	A	T	T	C	T	C
Haplotype 2	0.32	0.25	G	T	T	G	A	T	T	C	C	A	T	G	C	C	C	C	C
Haplotype 3	0.06	0.04	G	T	T	G	A	T	T	C	C	A	T	G	C	C	T	C	C
Haplotype 4	0.06	0.04	G	T	T	G	C	T	C	G	T	T	C	G	C	C	C	T	C

We used sections of synovial tissues for immunohistochemistry with antibodies to PADI4 and to citrulline. In each sample from an individual with rheumatoid arthritis, PADI4 protein was detected in the sublining (Fig. 3d). Citrullinated peptide was also detected in the sublining with a similar pattern (Fig. 3e). These results indicate that PADI4 protein and citrullinated peptides are localized in rheumatoid arthritis synovia.

Stability of two types of PADI4 mRNA

To investigate further the association between PADI4 alleles and rheumatoid arthritis, we tested whether SNPs in exons affect the stability of PADI4 mRNA. RNAs from the susceptible and non-susceptible alleles (Fig. 2c) were transcribed in vitro by modified RNase T1 selection assay[21]. Briefly, we mixed RNAs produced by in vitro transcription with extracts of HL-60 cells and observed the degradation of RNA by endogenous components of the HL-60 cell. Half-lives for susceptible and non-susceptible PADI4 mRNA were 11.6 min and 2.1 min, respectively. Susceptible mRNA was therefore significantly more stable than non-susceptible mRNA (after 5 min, $P = 0.038$; after 10 min, $P = 0.017$; Fig. 4). Based on this result, mRNA stability seems to depend on haplotype.

Relationship between SNP and antibody to citrullinated filaggrin

Citrullination in proteins is believed to create epitopes recognized by rheumatoid arthritis autoantibodies that not only represent the most specific serologic markers, but also appear early[22], even before clinical onset of rheumatoid arthritis. Citrullinated filaggrin has been used in clinical laboratory tests as a possible candidate for citrullinated

autoantigens[23]. We therefore examined the relationship between PADI4 haplotype and the presence of antibodies to citrullinated filaggrin in sera from individuals with rheumatoid arthritis. Individuals homozygous with respect to the susceptible haplotype were more likely to be positive (87%) for antibody to citrullinated filaggrin than the other two genotypes, for whom the positive fraction rate was 50% (Table 3). This tendency was tested using Fisher's exact test and was marginally significant (Table 4, $P = 0.038$).

DISCUSSION

A genome-wide association study to identify genes associated with rheumatoid arthritis is in progress in Japan using a high-throughput multiplex PCR-Invader assay[17,18]. Although the project has not yet been completed, one candidate locus has been identified in contig NT_034376.1. Previous sibling-pair linkage studies have also shown that this region is one of the three strongest susceptibility loci for rheumatoid arthritis[5,8]. This locus contains all four identified PADI genes, which encode calcium-dependent enzymes that catalyze the conversion of arginine to citrulline in peptides. This activity itself suggested that PADI genes may be involved in rheumatoid arthritis, and the antibodies are the most specific rheumatoid arthritis–specific antibodies identified[23–26]. Although several other genes with functional association to rheumatoid arthritis, including that encoding tumor necrosis factor receptor 2 (ref. 27), have been localized to 1p36, PADI genes were considered the most relevant for investigation owing to the rheumatoid arthritis specificity of the autoimmune response to citrullinated epitopes.
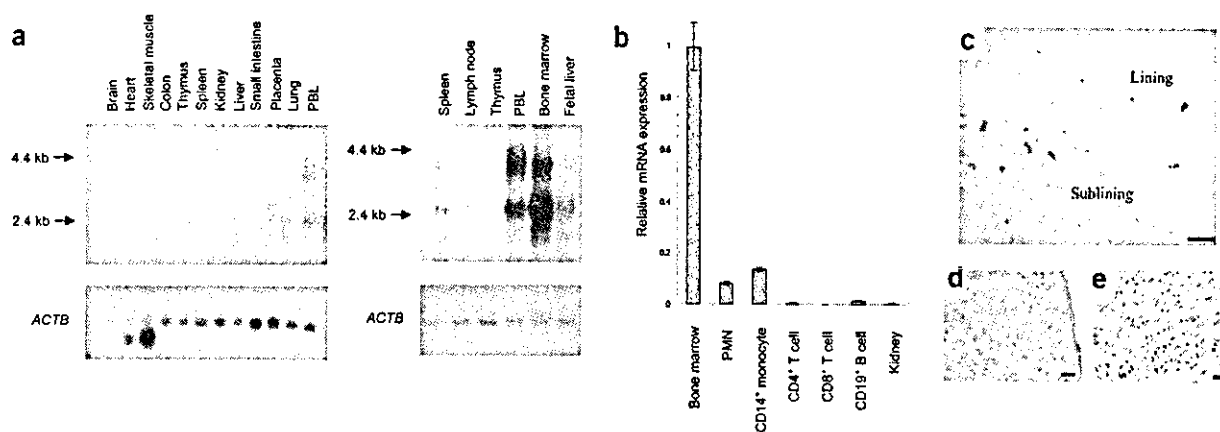

Figure 3 Expression of PADI4. (a) Expression of PADI4 mRNA in various normal human tissues. (b) Relative expression level of PADI4 mRNA in normal human tissues and cells. Values represent mean ± s.d. of data from triplicate wells. (c) Expression and distribution of PADI4 mRNA in rheumatoid arthritis synovial tissue as analyzed by in situ RT-PCR. PADI4 transcript (dark blue) was stained in sublining and lining. Immunohistochemistry showing expression patterns of PADI4 (d, red stain) and citrullinated peptides (e, red stain) in rheumatoid arthritis synovium. No non-rheumatoid arthritis tissue control was used. Scale bars: c, 250 μm; d, e, 100 μm.

We identified eight genes in contig NT_034376.1, including four PADI genes (Fig. 1a). We evaluated the strength of association with rheumatoid arthritis across the region by linkage disequilibrium mapping of 119 SNPs (Fig. 1b). The association in the region was definitive (P = 0.000008, OR = 1.97, 95% c.i. = 1.44–2.69) and was considered to originate in PADI4, rather than any other PADI gene (Fig. 1c). We observed a similar pattern of linkage disequilibrium in cases and controls, which is consistent with the association pattern (Fig. 1c) and provides additional support for PADI4 as the origin.

An OR of 1.97 suggests that the genetic contribution of PADI4 is not as strong as that of the HLA-DRB locus (OR = 2.60, 95% c.i. = 1.88–3.60; ref. 28) but is nonetheless considerable. The HLA-DRB locus has been estimated to explain less than or close to half of the total genetic contribution to rheumatoid arthritis, with the remainder attributed to multiple non-HLA genes[1]. We therefore expect that PADI4 is one of the primary non-HLA genes associated with rheumatoid arthritis. A genotypic risk ratio for PADI4 is 1.3 (ref. 29), and its population attributable risk is 17.4% (ref. 30), which seems reasonable for a gene associated with a complex genetic trait like rheumatoid arthritis. Furthermore, a locus with this degree of genetic contribution could be detectable in linkage studies, as was the case for microsatellite markers close to PADI4 in two linkage studies[5,8].

Northern-blot analysis indicated that PADI4 was highly expressed in bone marrow and peripheral blood leukocytes. Quantitative RT–PCR indicated that PADI4 mRNA is expressed in PMNs, which include neutrophils and the monocyte lineage, but is not expressed in lymphocytes. Previous reports have shown high PADI4 expression in neutrophils, eosinophils and monocytes[20,31]. PADI4 is therefore expressed in hematological tissues and cell types, which are known to be intimately involved in the pathogenesis of rheumatoid arthritis[32,33]. Although the importance of antigen-specific immune processes has been emphasized in the investigation of rheumatoid arthritis, the finding that myeloid leukocytes, rather than lymphocytes, are the predominant cell types in which PADI4 is expressed indicates that more investigation of the roles of myeloid lineages in rheumatoid arthritis is warranted.

We examined expression of PADI4 in synovial tissues of seven individuals with rheumatoid arthritis using in situ RT–PCR and immunohistochemistry. Both mRNA and protein were expressed in the sublining region, and both PADI4 protein and citrullinated peptide were localized in the sublining region. A previous study reported citrullinated α- and β-fibrin in sublining regions of fibroblast- and macrophage-like mononuclear cells of individuals with rheumatoid arthritis[14]. Peptides in synovial tissues, including fibrins, were proposed to be citrullinated by PADI4 extra- or intracellularly with subsequent secretion, behaving as autoantigens recognized by rheumatoid arthritis–specific antibodies. Lining regions contained PADI4 mRNA but no protein. The reason for this discrepancy is unclear. Collectively, these data suggest that citrullination by PADI4 occurs in the sublining of synovial tissues and that citrullinated peptides behave as antigens for rheumatoid arthritis–specific autoantibodies. Although the detection of PADI4 expression in rheumatoid arthritis synovial tissue without comparison to non-rheumatoid arthritis controls does not imply that expression and activity of PADI4 are specific to rheumatoid arthritis, its presence does support other findings that link rheumatoid arthritis and PADI4.
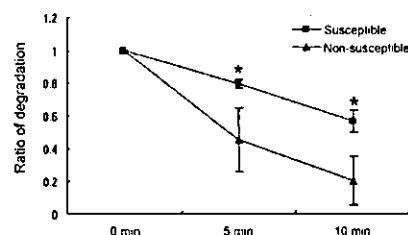

Figure 4 Stability of susceptible and non-susceptible transcripts of PADI4 mRNA measured as degradation rate. Differences were significant (*P < 0.05) after 5 min and 10 min of reaction time; Values represent mean ± s.d. of data from duplicate experiments.

To investigate the relationship between pathogenesis of rheumatoid arthritis and haplotypes comprising the four SNPs in PADI4 mRNA (Fig. 2c), we examined whether these SNPs affect PADI4 mRNA stability. The mRNA of the susceptible haplotype was more stable than that of the non-susceptible haplotype (Fig. 4). In previous studies, SNPs in mRNA or one-base deletions in coding regions have been associated with transcript stability[34,35]. The present result also suggests that SNPs in mRNA contribute to mRNA stability. Susceptible-haplotype mRNA probably accumulates to higher levels than non-susceptible mRNA, resulting in higher levels of PADI4 protein. Stable PADI4 mRNA may increase PADI4 proteins in synovial tissues, neutrophils and monocytes, increasing production of the citrullinated peptides that serve as autoantigens. Apart from stability of transcripts, evaluation of the effects of substitution of amino acids on the enzyme is important and further investigation should be directed at such analyses. Although SNPs in exons were systematically searched and the effect of coding SNPs analyzed in this report, involvement of other polymorphisms in non-coding regions is possible[35]. Further investigation in intron regions and other regulatory areas is therefore desirable.

The relationship between PADI4 and rheumatoid arthritis is further supported by the fact that the positive fraction of antibodies to citrullinated peptides was significantly higher in individuals homozygous with respect to the susceptible haplotype than in those of other genotypes (P = 0.038, Table 4). The present study yielded statistically significant results only in comparing susceptible homozygotes with others and not in comparing non-susceptible homozygotes with others. The absence of a significant difference in the latter comparison might be due to the small number of samples or the mixture of individuals with positive results irrelevant to rheumatoid arthritis–related PADI activity, as should be observed in healthy controls using a test with sensitivity of 75.6%. Previous reports that antibodies to citrullinated peptide are specific to rheumatoid arthritis and are detectable in the early phases of the disease[36] suggest that citrullination by

Table 3 Distribution of individuals of each genotype that were positive for antibody to citrullinated filaggrin

SNP genotype			
Antibody to citrullinated filaggrin	Susceptible homozygotes	Heterozygotes	Non-susceptible homozygotes
Positive	26 (30%)	40 (45%)	22 (25%)
Negative	4 (11%)	20 (57%)	11 (11%)
Positive fraction	0.87	0.50	0.50

— 290 —

ARTICLES

Table 4 Association test between genotype and antibody positivity

Comparison pattern	P value*
Susceptible homozygotes versus others	0.038
Non-susceptible homozygotes versus others	0.468

*P value was calculated by Fisher's exact test (two-tailed).

PADI4 should be closely linked to onset of rheumatoid arthritis or might represent a disease-triggering event in itself.

To investigate the precise role of PADI4 in rheumatoid arthritis, we evaluated the mouse homolog of PADI4, Padi4, in a collagen-induced arthritis (CIA) mouse model. Expression of Padi4 was quantified (**Supplementary Fig. 1** online). We induced expression of Padi4 in inflamed synovial tissues and spleen in mice with CIA. In humans, genotype with respect to PADI4 was associated with rheumatoid arthritis, presence of PADI4 in affected joints was detected and antibody to citrullinated peptide was detected in sera. In mice, expression of Padi4 increased with appearance of CIA, but antibody to citrullinated peptide was not detected in sera (data not shown). The primary difference between human rheumatoid arthritis and mouse CIA is that the former is characterized by breakdown of self-tolerance and continuity of destructive arthritis with accompanying autoimmune phenomena to various autoantigens including antibody to citrullinated proteins, whereas the latter shares the inflammatory component related to immune response to collagen type II with rheumatoid arthritis, but specificity of its immunoreaction is higher and breakage of tolerance to citrullinated antigens does not seem to be involved. Given the results of the present study, we consider citrullination by PADI4 or Padi4 as one of the processes in early phase arthritis, and that, in human rheumatoid arthritis, immunological tolerance breaks down somehow with the appearance of autoantibody recognizing citrullinated peptide, followed by the autoimmune disease process characterized for rheumatoid arthritis. In mouse CIA, however, expression of Padi4 increases with a probable increase in citrullination of self-peptides, but tolerance to citrullinated-antigens does not seem to break. Even with these differences in mechanisms between human rheumatoid arthritis and mouse CIA, further investigation of PADI4 in human rheumatoid arthritis and Padi4 in the mouse model seems warranted.

In conclusion, we identified PADI4 as a susceptibility gene for rheumatoid arthritis using a case-control study with SNPs. The present findings imply that the rheumatoid arthritis susceptibility haplotype in PADI4 produces a more stable transcript and is associated with higher levels of antibody to citrullinated peptide in sera of individuals with rheumatoid arthritis. Given the polygenic nature of rheumatoid arthritis, this independent susceptibility gene could have a most important role in rheumatoid arthritis pathogenesis by increasing citrullination of proteins in rheumatoid arthritis synovial tissues, leading, in a cytokine-rich milieu, to a break in tolerance to citrullinated peptides processed and presented in the appropriate HLA context.

METHODS

Subjects with rheumatoid arthritis and unaffected subjects. We recruited a total of 830 individuals affected with rheumatoid arthritis and 736 unaffected controls for collection of genomic DNA and sera through several medical institutes in Japan. We sampled pathological joint synovial tissues from seven individuals with rheumatoid arthritis who underwent arthroplasty surgery. All rheumatoid arthritis cases met the revised criteria of the American College of Rheumatology for rheumatoid arthritis[37]. The mean age of the 830 case individuals with rheumatoid arthritis was 64.3 y (range, 28–92 y). Most case subjects were female (83.7%), and 75% were positive for rheumatoid factor. Control subjects comprised 736 individuals from the general population, 57.4% females, with mean age of 48.6 y (range, 3–92 y). We obtained informed consent from each subject, with parental authority in the case of minors, as approved by the ethical committee of the SNP Research Center of The Institute of Physical and Chemical Research (RIKEN).

SNPs. We identified four SNPs in exons of PADI4 and 14 SNPs in LOC148695 by direct sequencing of DNA from 48 case individuals. We selected the other 101 SNPs, which were located in genes (promoter, exon and intron) in NT_034376.1 (gi: 22043311) from the JST database.

Genotyping. We extracted genomic DNA from peripheral blood leukocytes using standard protocols[17]. We genotyped SNPs using the Invader assay, TaqMan assay or direct sequencing. For Invader assay, we amplified DNA with PCR primers designed to include one or more SNPs, as previously described[18,38]. Third Wave Technologies designed probe sets for each locus. In TaqMan assay, we carried out PCR using TaqMan Universal Master Mix (Applied Biosystems), 8 ng DNA, 1 μM of each primer and 200 nM of probe in 15-μl reactions. Each 96-well plate contained 94 samples of unknown genotype and 2 no-DNA control samples. Thermal cycle conditions were 50 °C for 2 min, 95 °C for 10 min, 50 cycles of 92 °C for 15 s and 58 °C for 1 min. Thermal cycling was done on an ABI PRISM 7700 Sequence Detector Systems (Applied Biosystems). We undertook direct sequencing of PCR products using ABI3700 capillary sequencers (Applied Biosystems) according to standard procedures.

Northern-blot hybridization. We hybridized human multiple tissue northern (MTN) blots (Clontech) with a PADI4 probe labeled with digoxigenin. We generated digoxigenin-labeled PADI4 probes using a PCR digoxigenin probe synthesis kit (Roche Diagnostics) according to the manufacturer's instructions, using the primers to generate a 335-bp product. Hybridization and detection were also done according to the manufacturer's instructions. Blots were stripped of probe and re-hybridized with a cDNA probe for ACTB (Roche Diagnostics) to assess RNA loading. Primer sequences are available on request.

RNA extraction and cDNA synthesis. We separated PMNs using Mono-Poly resolving solution (Dainippon Pharmaceuticals) and extracted RNA from PMNs using ISOGEN (Nippon Gene). We stored the resulting RNA at –80 °C until use. We quantified RNAs of other normal tissues using Premium total RNA (Clontech). We reverse-transcribed total RNA (1 μg) using a First Strand cDNA synthesis kit (Amersham Pharmacia) according to the manufacturer's instructions.

Quantification of PADI4 expression by real-time RT–PCR. We carried out real-time PCR on the ABI PRISM 7000 (Applied Biosystems) using QuantiTect SYBR Green PCR (QIAGEN) according to the manufacturer's instructions. Each oligonucleotide primer set was added to a final concentration of 0.3–0.5 μM for ACTB (product size, 219 bp) and PADI4 (product size, 207 bp). We generated a standard curve from data of amplification of PADI4 primers using a dilution series of bone marrow mRNA as templates and normalized to ACTB. Primer sequences are available on request.

In situ RT–PCR. We carried out one-step *in situ* RT–PCR by adding Pro STAR HF (Stratagene), and reactions were done using an Omnislide thermal cycler (Hybaid) as follows: (i) 42 °C for 30 min; (ii) 94 °C for 2 min, 55 °C for 45 s and 68 °C for 2 min; and (iii) 25 cycles at 94 °C for 45 s, 55 °C for 45 s and 68 °C for 2 min. Reactions were maintained at 4 °C after amplification. After PCR, we washed slides twice with Tris-buffered saline for 5 min. Specific primers amplified their specific target sequences, yielding a 335-bp product.

We generated digoxigenin-labeled internal probes by PCR using the PCR digoxigenin probe synthesis kit according to the manufacturer's specification with minor modifications. We added primers to a final concentration of 0.34 μM. We covered slides with pre-hybridization solution at 37 °C for 1 h. After pre-hybridization, we replaced pre-hybridization solution with hybridization

solution containing probe. Probes were denatured at 94 °C for 5 min. We carried out hybridization for 12 h at 37 °C. After washing, we visualized incorporated PCR fragments using a digoxigenin detection kit (Roche Diagnostics).

Controls included several different samples, substituting water for primer in the PCR reaction, omitting reverse transcription in the case of mRNA and omitting probe in hybridization solutions (X.C. *et al.*, manuscript in preparation). Primer sequences are available on request.

Preparation and purification of antiserum against PADI4. We synthesized *PADI4*-derived peptides (Sp-PAD1: PAKKK STGSS TWP-Cys), purified and immunized in rabbits (Kitayama-Labes, Nagano, Japan). We purified antiserum by affinity chromatography on a histidine-tagged PADI4 column (Bio-Gate). We confirmed specificity of purified polyclonal antibody to PADI4 with western blotting using a transient expression system in the HEK293 cell line (data not shown).

Immunohistochemistry. We incubated paraffin sections of synovial tissues at 4 °C for 12 h with rabbit polyclonal antibody to PADI4 or with rabbit antibody to citrulline (Biogenesis), diluted at 1:1,000. We washed and incubated sections at room temperature with Simple Stain MAX-PO (Nichirei) for 30 min and then added Simple Stain AEC (Nichirei). We incubated sections for 5–20 min until the reaction was obviously visible under light microscopy. All sections were counterstained with hematoxylin. In all cases, negative controls omitted the specific antibody and used normal mouse and rabbit antiserum.

Measurement of antibody to citrullinated filaggrin. We measured levels of antibody to citrullinated filaggrin using an ELISA kit (MBL) according to the manufacturer's instructions. Sensitivity was 75.6% and specificity was 83.2% for testing subjects with and without rheumatoid arthritis n clinical settings at a cutoff level of 9 (K. Suzuki *et al.*, manuscript accepted).

In vitro RNA stability assay. We amplified genes encoding two *PADI4* variants by PCR from cDNAs that were synthesized using a first-strand cDNA synthesis kit (Amersham Pharmacia) with bone marrow total RNA (Clontech). We then cloned these genes into the pDONR201 vector (Invitrogen). We also constructed the cDNA into pDEST14 (Invitrogen), which has a T7 promoter, and sequenced both strands of the resulting expression vector. Vectors were digested using ClaI, and both types of *PADI4* were expressed using RiboMax Large Scale RNA Production System-T7 (Promega) and purified according to the manufacturer's instructions. To prepare whole-cell extract, we washed HL-60 cells in phosphate-buffered saline and re-suspended them in extraction buffer (0.5% Nonidet P-40; 20 mM HEPES buffer, pH 8.0; 20% glycerol (v/v); 400 mM NaCl, 0.5 mM dithiothreitol; 0.2 mM EDTA and 1% protease inhibitor cocktail (Nacalai)). After incubation on ice for 30 min and microcentrifugation at 4 °C, we transferred supernatants to new tubes and stored them at –80 °C until use.

We mixed and incubated each 5 µg of synthesized RNA and diluted whole-cell extract (1:1,000) at room temperature. The reaction was stopped with the addition of formamide dye, and the samples were then heated at 68 °C. After the reaction, we detected RNA using northern-blot hybridization. We scanned results on a DocuCentre Color 500cp (Fuji-Xerox) and measured signal intensities of full-length RNAs using Adobe Photoshop 6.0.

Statistical analysis. We estimated haplotype frequencies using the expectation-maximization algorithm[39]. We calculated linkage disequilibrium index, Δ (ref. 40), and drew Figure 1c with an application created by our group with the assistance of Excel (Microsoft). Associations between phenotypes were estimated by χ^2 test. Antibody to citrullinated filaggrin titer and genotypes were tested using Fisher's exact test on Statistica software (StatSoft), and mRNA stability data and quantitative RT-PCR data were tested using Student's *t*-test.

URLs. The National Center for Biotechnology Information can be found at http://www.ncbi.nlm.nih.gov/. The International RH Mapping Consortium can be found at http://www.ncbi.nlm.nih.gov/genemap99/. The expectation-maximization program can be found at http://linkage.rockefeller.edu/ott/eh.htm.

GenBank accession numbers. *PADI4*, NM_012387; *LOC148695*, XM_088976; *Padi4*, NM_011061.

1. Seldin, M.F., Amos, C.I., Ward, R. & Gregersen, P.K. The genetics revolution and the assault on rheumatoid arthritis. *Arthritis Rheum.* 42, 1071–1079 (1999).
2. Gregersen, P.K., Silver, J. & Winchester, R.J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.* 30, 1205–1213 (1987).
3. Nepom, G.T. Major histocompatibility complex-directed susceptibility to rheumatoid arthritis. *Adv. Immunol.* 68, 315–332 (1998).
4. Weyand, C.M. & Goronzy, J.J. Association of MHC and rheumatoid arthritis. HLA polymorphisms in phenotypic variants of rheumatoid arthritis. *Arthritis Res.* 2, 212–216 (2000).
5. Cornelis, F. *et al.* New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc. Natl. Acad. Sci. USA* 95, 10746–10750 (1998).
6. Jawaheer, D. *et al.* A genome-wide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am. J. Hum. Genet.* 68, 927–936 (2001).
7. MacKay, K. *et al.* Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum.* 46, 632–639 (2002).
8. Shiozawa, S. *et al.* Identification of the gene loci that predispose to rheumatoid arthritis. *Int. Immunol.* 10, 1891–1895 (1998).
9. Schellekens, G.A., de Jong, B.A., van den Hoogen, F.H., van de Putte, L.B. & van Venrooij, W.J. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *J. Clin. Invest.* 101, 273–281 (1998).
10. Menard, H.A., Lapointe, E., Rochdi, M.D. & Zhou, Z.J. Insights into rheumatoid arthritis derived from the Sa immune system. *Arthritis Res.* 2, 429–432 (2000).
11. Schellekens, G.A. *et al.* The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis Rheum.* 43, 155–163 (2000).
12. Nogueira, L. *et al.* Performance of two ELISAs for antifilaggrin autoantibodies, using either affinity purified or deiminated recombinant human filaggrin, in the diagnosis of rheumatoid arthritis. *Ann. Rheum. Dis.* 60, 882–887 (2001).
13. Goldbach-Mansky, R. *et al.* Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. *Arthritis Res.* 2, 236–243 (2000).
14. Masson-Bessiere, C. *et al.* The major synovial targets of the rheumatoid arthritis-specific antifilaggrin autoantibodies are deiminated forms of the α- and β-chains of fibrin. *J. Immunol.* 166, 4177–4184 (2001).
15. Baeten, D. *et al.* Specific presence of intracellular citrullinated proteins in rheumatoid arthritis synovium: relevance to antifilaggrin autoantibodies. *Arthritis Rheum.* 44, 2255–2262 (2001).
16. Zhou, Z. & Menard, H.A. Autoantigenic posttranslational modifications of proteins: does it apply to rheumatoid arthritis? *Curr. Opin. Rheumatol.* 14, 250–253 (2002).
17. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* 32, 650–654 (2002).
18. Ohnishi, Y. *et al.* A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* 46, 471–477 (2001).
19. McIntyre, L.M., Martin, E.R., Simonsen, K.L. & Kaplan, N.L. Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. *Genet. Epidemiol.* 19, 18–29 (2000).
20. Nakashima, K. *et al.* Molecular characterization of peptidylarginine deiminase in HL-60 cells induced by retinoic acid and 1alpha,25-dihydroxyvitamin D(3). *J. Biol. Chem.* 274, 27786–27792 (1999).
21. Cuadrado, A. *et al.* HuD binds to three AU-rich sequences in the 3' UTR of neuroserpin mRNA and promotes the accumulation of neuroserpin mRNA and protein. *Nucleic Acids Res.* 30, 2202–2211 (2002).
22. Vincent, C. *et al.* Detection of antibodies to deiminated recombinant rat filaggrin by enzyme-linked immunosorbent assay: a highly effective test for the diagnosis of rheumatoid arthritis. *Arthritis Rheum.* 46, 2051–2058 (2002).
23. van Venrooij, W.J. & Pruijn, G.J. Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. *Arthritis Res.* 2, 249–251 (2000).
24. Vincent, C. *et al.* High diagnostic value in rheumatoid arthritis of antibodies to the stratum corneum of rat oesophagus epithelium, so-called 'antikeratin antibodies'. *Ann. Rheum. Dis.* 48, 712–722 (1989).
25. Gomes-Daudrix, V. *et al.* Immunoblotting detection of so-called 'antikeratin antibodies': a new assay for the diagnosis of rheumatoid arthritis. *Ann. Rheum. Dis.* 53, 735–742 (1994).

— 292 —

26. Vincent, C. *et al.* Immunoblotting detection of autoantibodies to human epidermis filaggrin: a new diagnostic test for rheumatoid arthritis. *J. Rheumatol.* **25**, 838–846 (1998).

27. Shibue, T. *et al.* Tumor necrosis factor α 5′-flanking region, tumor necrosis factor receptor II, and HLA-DRB1 polymorphisms in Japanese patients with rheumatoid arthritis. *Arthritis Rheum.* **43**, 753–757 (2000).

28. de Vries, N., Tijssen, H., van Riel, P.L. & van de Putte, L.B. Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67-74 of the HLA-DRB1 molecule. *Arthritis Rheum.* **46**, 921–928 (2002).

29. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).

30. Schildkraut, J.M. Examining complex genetic interactions. in *Approach to Gene Mapping in Complex Human Diseases* (eds. J.L. Haines and M.A. Pericak-Vance) 379–410 (Wiley-Liss, New York, 1998).

31. Asaga, H., Nakashima, K., Senshu, T., Ishigami, A. & Yamada, M. Immunocytochemical localization of peptidylarginine deiminase in human eosinophils and neutrophils. *J. Leukoc. Biol.* **70**, 46–51 (2001).

32. Pillinger, M.H. & Abramson, S.B. The neutrophil in rheumatoid arthritis. *Rheum. Dis. Clin. North Am.* **21**, 691–714 (1995).

33. Hirano, T. Revival of the autoantibody model in rheumatoid arthritis. *Nat. Immunol.* **3**, 342–344 (2002).

34. Zhou, X. *et al.* Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. *Arthritis Rheum.* **46**, 2990–2999 (2002).

35. Yang, T., McNally, B.A., Ferrone, S., Liu, Y. & Zheng, P. A single nucleotide deletion leads to rapid degradation of TAP-1 mRNA in a melanoma cell line. *J. Biol. Chem.* **278**, 15291–15296 (2003).

36. Jansen, A.L. *et al.* Rheumatoid factor and antibodies to cyclic citrullinated Peptide differentiate rheumatoid arthritis from undifferentiated polyarthritis in patients with early arthritis. *J. Rheumatol.* **29**, 2074–2076 (2002).

37. Arnett, F.C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).

38. Hirakawa, M. *et al.* JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res.* **30**, 158–162 (2002).

39. Ott, J. Counting methods (EM algorithm) in human pedigree analysis: linkage and segregation analysis. *Ann. Hum. Genet.* **40**, 443–454 (1977).

40. Devlin, B. & Risch, N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**, 311–322 (1995).

— 293 —

Reconstitution of CD8$^+$ T Cells by Retroviral Transfer of the TCR αβ-Chain Genes Isolated from a Clonally Expanded P815-Infiltrating Lymphocyte[1]

Hiroyuki Tahara,* Keishi Fujio,* Yasuto Araki,* Keigo Setoguchi,* Yoshikata Misaki,* Toshio Kitamura,[†] and Kazuhiko Yamamoto[2]*

Gene transfer of TCR αβ-chains into T cells may be a promising strategy for providing valuable T lymphocytes in the treatment of tumors and other immune-mediated disorders. We report in this study the reconstitution of CD8$^+$ T cells by transfer of TCR αβ-chain genes derived from an infiltrating T cell into P815. Analysis of the clonal expansion and Vβ subfamily usage of CD8$^+$ TIL in the tumor sites demonstrated that T cells using Vβ10 efficiently infiltrated and expanded clonally. The TCR α- and β-chain sequences derived from a tumor-infiltrating CD8$^+$/Vβ10$^+$ single T cell clone (P09-2C clone) were simultaneously determined by the RT-PCR/single-strand conformational polymorphism method and the single-cell PCR method. When P09-2C TCR αβ-chain genes were retrovirally introduced into CD8$^+$ T cells, the reconstituted T cells positively lysed the P815 tumor cells, but not the A20, EL4, or YAC-1 cells, in vitro. In addition, the CTL activity was blocked by the anti-H2L^d mAb. Furthermore, T cells containing both TCR α- and β-chains, but not TCR β-chain alone, accumulated at the tumor-inoculated site when the reconstituted CD8$^+$ T cells were adoptively transferred to tumor-bearing nude mice. These findings suggest that it is possible to reconstitute functional tumor-specific CD8$^+$ T cells by transfer of TCR αβ-chain genes derived from TIL, and that such T cells might be useful as cytotoxic effector cells or as a vehicle for delivering therapeutic agents. *The Journal of Immunology*, 2003, 171: 2154–2160.

R ecently, novel strategies have been introduced for the study of immune responses against various tumors. An increasing number of tumor Ags have been found by screening for gene products differentially expressed in tumors as opposed to normal tissues and by testing for antigenicity (1). These Ags, including the form of peptides, proteins, DNA, RNA, or viral vectors, are currently used in clinical trials. Effective T cell activation depends on the presentation of these Ags by APCs, such as dendritic cells. Vaccination trials using hybrids of tumor Ags and dendritic cells have demonstrated their beneficial effect for some tumors (2–4), but these strategies do not appear to be reliably sufficient for the real patient treatments.

The CD8$^+$ CTL play a significant role in Ag-specific immune responses to tumors or pathogens by their cytolytic activity. Tumor-reactive CD8$^+$ T lymphocytes can often be isolated from tumors in mice or patients, and these have been used to study Ag specificity (5), frequency of T cell repertoire (6, 7), tumor Ag gene cloning (8, 9), and experimental adoptive immunotherapy (10–12). Adoptive transfer of tumor-specific CTL has resulted in regression or eradication of murine tumors (10, 13) and reduction of pulmonary metastases (14). In contrast, functional studies of tu-

mor-infiltrating lymphocytes (TIL)[3] are difficult because of their weak activity, although many attempts have been made to analyze the immune responses by freshly isolated TIL. It has also been demonstrated that TIL expanded ex vivo and then adoptively transferred to melanoma patients with IL-2 resulted in objective responses in 34% of melanoma patients (15). This observation indicated that TIL may have potent antitumor activities.

It is generally believed that each T cell has a distinct clonotype of TCR that is responsible for the Ag-specific T cell response. The complementarity-determining regions 3 (CDR3) of TCR seem to play key roles in Ag recognition (16). We previously established a method for analyzing T cell clonality by the RT-PCR/single-strand conformational polymorphism (SSCP) method (17). This method detects nucleotide changes of CDR3 of clonally expanded T cells in vivo. Using this method, we have demonstrated that in patients with solid tumors, T cell oligoclonal expansion was found to extend from the original tumor to the draining lymph nodes and to PBLs during metastasis (18). We have also shown that ovarian tumor-infiltrating T cells exhibited clonal expansion and that such T cells were specific for autologous tumors (19), suggesting that the clonally expanded T cells in the tumor site recognize a tumor Ag in vivo. These findings indicate that the information of specific TCR expressed in the tumor site may make it possible to reconstitute functional tumor-specific CTL.

Several studies have been made on the reconstitution of functional T cells by TCR αβ-chain gene transfer, but most of these studies have examined the reconstitution of TCRs into T cell hybridomas and T cell lines (20–24). We have recently reported a functional reconstitution of peripheral T lymphocytes by retroviral

*Department of Allergy and Rheumatology, Graduate School of Medicine, and †Division of Cellular Therapy, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan

[2] Address correspondence and reprint requests to Dr. Kazuhiko Yamamoto, Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8655, Japan. E-mail address: yamamoto-tky@umin.ac.jp

[3] Abbreviations used in this paper: TIL, tumor-infiltrating lymphocyte; CDR, complementarity-determining region; GFP, green fluorescent protein; IRES, internal ribosomal entry site; RT, reverse transcription; SSCP, single-strand conformational polymorphism; T1, tumor 1; T2, tumor 2.

transfer of TCR $\alpha\beta$-chain cDNAs for restoration of immunity (25). Namely, reconstituted CD4$^+$ T cells with OVA-specific TCR strongly responded to OVA peptide in the in vitro culture and induced an Ag-specific delayed-type hypersensitivity in response to the OVA peptide challenge in vivo. Moreover, other groups have demonstrated that peripheral T lymphocytes may be reconstituted by Ag-specific TCR gene transfer (26–29).

In this study, we analyzed the TIL derived from the P815 tumor site and the TCR α- and β-chain DNA sequences of clonally expanded CD8$^+$ single TIL by the RT-PCR/SSCP method and the single-cell PCR method. We also examined the cytotoxic response of the TCR $\alpha\beta$-chain-introduced CD8$^+$ T cells to the original tumor in vitro and the recruitment of these T cells to the P815 tumor site in vivo.

Materials and Methods

Mice

DBA/2J and nude mice (BALB/c *nu/nu*) were obtained from Japan SLC (Shizuoka, Japan). All the mice used were female and were used at 7–8 wk of age.

Cytokines and Abs

Recombinant murine IL-2 and IL-12 were obtained from R&D Systems (Minneapolis, MN). Fc block (anti-CD16/CD32), PE anti-CD8, PE anti-Vβ10, FITC anti-TCR Vβ chain (H57-597), and FITC anti-Vβ2, 4, 7, 10, 13, and 14 mAb were obtained from BD PharMingen (San Diego, CA). Anti-H2K^d, anti-H2D^d, anti-H2L^d, and control IgG were obtained from Cedarlane (Ontario, Canada).

Cell culture

Mastocytoma P815 cells (30) were cultured in complete medium consisting of RPMI 1640 medium supplemented with 10% FCS, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, and 0.05 mM 2-ME. Splenocytes were cultured in complete medium supplemented with 50 pg/ml murine IL-2, 2 ng/ml murine IL-12, and 10 μg/ml Con A for 48 h before retroviral infection.

Isolation of CD8$^+$ TILs and flow cytometric analysis

DBA/2 mice were injected s.c. with 1×10^6 tumor cells. After 14–18 days, tumors were resected and each was divided into three pieces (Fig. 2A). One piece was minced to yield 1- to 2-mm pieces. To release the tumor cells and TIL, the tumor pieces were incubated in a mixture of 1 mg/ml type IV collagenase (Sigma-Aldrich, St. Louis, MO) and 20 μg/ml DNase (Sigma-Aldrich) in complete medium for 90 min at 37°C. The cell suspension was strained through nylon mesh and washed with PBS. The CD8$^+$ TIL were isolated from the cell suspension by the treatment of Fc block and PE-labeled anti-CD8 mAb, followed by collection with anti-PE-conjugated paramagnetic beads and a MACS LS separation column, according to the manufacturer's instructions (Miltenyi Biotech, Bergisch Glandbach, Germany). For flow cytometric analysis, tumors were resected from four P815-bearing mice, and the CD8$^+$ TIL were isolated, as described above. The CD8$^+$ TIL batches were mixed and stained with FITC-conjugated anti-TCR Vβ chain or Vβ mAbs.

Oligonucleotides

Reverse-transcription (RT) reactions were performed using RT primers (CαRT, 5'-AGC TTG TCT GGT TGC TCC AG-3' and CβRT, 5'-TGT GCC AGA AGG TAG CAG AG-3'). The TCR α-chains were amplified using TCRα first primer set (AL3N, 5'-TCT TCA GAA TTC TTT TTT TTT TTT TTT TTT TTT TTT-3' and Cα first, 5'-GTT TTG TCA GTG ATG AAC GT-3'), TCRα second primer set (AL3N and Cα second, 5'-TCG GCA CAT TGA TTT GGG AG-3'), and TCRα third primer set (AL3N and Cα third, 5'-AAG TCG GTG AAC AGG CAG AG-3'). The TCR β-chains were amplified using TCRβ first primer set (Vβ10–1st, 5'-GCA AGA CTC TAA GAA ATT GC-3' and Cβ first, 5'-TGG ACT TCC TTG CCA TTC AC-3'), TCRβ second primer set (Vβ10–second, 5'-CTC ATT GTA AAC GAA ACA GT -3' and Cβ second, 5'-TTC ACC CAC CAG CTC AGC TC-3'), and TCRβ third primer set (Vβ10–third, 5'-AAT CAA GTC TGT AGA GCC GG-3' and Cβ third, 5'-GGC TCA AAC AAG GAG ACC TTG-3').

Single-cell sorting and RT-PCR

The CD8$^+$ TIL were stained with FITC-conjugated anti-Vβ10. The CD8$^+$/Vβ10$^+$ cells were sorted at a ratio of one cell/well using an automatic cell dispensing unit driven by the FACSVantage and Clone-Cyt software (BD Biosciences, Franklin Lakes, NJ). Each cell was sorted into 20 μl of RT reaction mixture (10 nM CαRT primer, 10 nM CβRT primer, 1× RT reaction buffer, 100 μM each dNTP (Takara, Ohtsu, Japan), 0.5% Nonidet P-40 (Boehringer Mannheim, Mannheim, Germany), and 0.5 U/μl RNasin (Promega, Madison, WI)) in a 96-well microtiter plate. Immediately, 20 U/μl SUPERSCRIPT II (Life Technologies, Gaithersburg, MD) reagent was added to each well, and the plate was held at 37°C for 90 min. After the reaction mixture received heat inactivation for 10 min at 65°C, an equal volume of TdT solution (2× TdTase reaction buffer, 2.5 mM dATP (Amersham Pharmacia Biotech, Piscataway, NJ), and 0.5 U/μl TdT (Life Technologies, Rockville, MD)) was added to each well, and the plate was incubated for 15 min at 37°C (31). From the single-cell RT reaction mixtures, 2 μl cDNA was added to 23 μl of first PCR premix (1.6 pmol/μl each first primer, 200 mM each dNTP, and 0.25 U/μl KOD-plus-*Taq* polymerase (Toyobo, Osaka, Japan)) and amplified by a 25-cycle program (95°C for 1 min, 52°C for 1 min, 72°C for 2 min). Two microliters of first PCR products were used for the second PCR (30 cycles of 95°C for 1 min, 54°C for 1 min, 72°C for 2 min), using the second PCR premix (1.6 pmol/μl each second primer, 200 mM each dNTP, and 0.25 U/μl *Taq* polymerase (Promega)). Then 2 μl of the second PCR products was used for further amplification reaction (35 cycles of 95°C for 1 min, 54°C for 1 min, and 72°C for 2 min), using the third PCR premix (1.6 pmol/μl each third primer, 200 mM each dNTP, and 0.25 U/μl *Taq* polymerase).

DNA sequencing

For DNA sequencing, TCR β-chain bands on a SSCP gel were cut out and amplified, as described previously (18). Amplified cDNA was purified using SUPREC-01 membrane filter (Takara) and cloned into the pGEM T vector (Promega). Plasmid DNA from positive colonies was sequenced with DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) by an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). For the TCR α-chain, the amplified PCR products from a single cell were sequenced in the same manner as the TCR β-chain.

RNA extraction and cDNA synthesis

Solid tumors were resected from mice that had been injected s.c. with 1×10^6 P815 tumor cells and divided. Total RNA from the tumor pieces was isolated by the acid guanidinium thiocyanate-phenol-chloroform extraction method (15) using ISOGEN (Wako, Tokyo, Japan). Total RNA (20 μg) was converted into cDNA with random primers (Life Technologies) and SUPERSCRIPT II reverse transcriptase.

SSCP

The SSCP study was performed, as described previously (17, 18). In brief, the synthesized cDNA was amplified by PCR with a pair of Vβ1 to Vβ19 primers and a Cβ common primer. The amplified DNA was electrophoresed on a nondenaturing 4% polyacrylamide gel. After transfer onto a nylon membrane, the cDNA was hybridized with a biotinylated internal common Cβ oligonucleotide probe and visualized by subsequent incubations with streptavidin, biotinylated alkaline phosphatase, and a chemiluminescent substrate system (Phototope-Star Chemiluminescent Detection Kit; New England Biolabs, Beverly, MA).

Construction of TCR expression vectors

Because it is difficult to isolate the full-length P09-2C cDNA from a single cell, we first cloned the full-length cDNA encoding the Vα3 and Vβ10 TCR that has random CDR3 regions. Then each of the random CDR3 genes was exchanged to the P09-2C CDR3 genes using site-directed mutagenesis by PCR (32). The full-length TCR α- or β-chains were inserted into the retrovirus vector pMX (33). For cell transfer experiments, the full-length TCR α-chain was inserted into the pMX-IRES (internal ribosomal entry site)/GFP (green fluorescent protein) vector. The expression plasmids were transfected to PLAT-E cells, and the supernatants were collected, as previously described (34). An efficient packaging cell line, PLAT-E, can produce retroviruses with a titer of $\sim 1 \times 10^7$/ml (20).

Retrovirus infection to splenocytes

The mixture of both viral supernatants was placed on nontissue culture 24-well plates coated with RetroNectin and centrifuged for 3 h at 700 × g at room temperature. This step was repeated three times, and then splenocytes were placed into the plates and cultured for 24 h.

2156 RECONSTITUTION OF CD8⁺ T CELLS BY RETROVIRAL TRANSFER OF TCR

Cytotoxicity assay

Cytotoxicity was measured with a standard ^{51}Cr release assay. Plasmid-transduced splenocytes were used as effector cells, and the tumor cells were used as targets. The lysis of target cells was determined by the release of ^{51}Cr after 4 h of coincubation with effector cells. In the case of Fig. 4, *C* and *D*, data points are represented as Δ% specific lysis. Δ% specific lysis = (% specific lysis of TCRαβ-introduced T cells) − (% specific lysis of mock-introduced T cells).

Blocking experiments

Anti-H2K^d, anti-H2D^d, or anti-H2L^d mAb was added to the mixture of effector and target cells in 200 μl at an optimal concentration, according to the manufacturer's instruction, and then incubated for 4 h at 37°C.

Cell transfer experiments

TCRαβ (pMX-TCRα-IRES/GFP + pMX-TCRβ), TCRβ (pMX-IRES/GFP + pMX-TCRβ), or mock expression vectors (pMX- IRES/GFP + pMX) were introduced into splenocytes from DBA/2 mice. For cell transfer, CD8⁺ T cells were isolated from the infected splenocytes by negative selection using MACS LD column. Nude mice were injected s.c. with 1 × 10⁶ tumor cells and i.v. with 2 × 10⁶ infected CD8⁺ T cells. After 12 days, CD8⁺ TIL were isolated, stained with PE anti-Vβ10, and subsequently analyzed by flow cytometry.

Statistical analysis

The statistical significance of the differences was analyzed with Dunnett's *t* test or least significant difference multiple comparison test using SPSS software.

Results

Analysis of P815-infiltrating T cells by SSCP

The clonal expansion of Vβ subfamilies among infiltrating T cells in the P815 tumor site was initially analyzed by the RT-PCR/SSCP method (17, 18). Because each CDR3 sequence demonstrates unique mobility due to its unique single-strand conformation, identical mobility of the amplified CDR3 region on a gel indicates that the clones are identical (35). Therefore, a single T cell clone makes a band, which means an expanded clone.

Mice were injected s.c. with 1 × 10⁶ tumor cells. After 14–18 days, the tumors were resected and divided, and cDNA was subsequently recovered from them and amplified by PCR with each Vβ primer and a Cβ common primer. We analyzed each amplified Vβ CDR3 region by SSCP and found clonal expansion in several Vβ chains (data not shown). The clonal expansions of Vβ1, 2, 4, 7, 10, 13, and 14 were always detected, and multiple identical bands were observed in different samples called tumor 1 (T1) and tumor 2 (T2) of a mouse, indicating that homogenous Ag-specific immune responses existed in the different areas of the tumor site (Fig. 1). In contrast, peripheral (spleen) repertoires exhibited a smear pattern (data not shown), as described previously (17), because the CDR3 regions of T cells are diverse. Moreover, we performed flow cytometric analysis to investigate Vβ usage of CD8⁺ TIL. The P815 solid tumors were resected from four mice, and CD8⁺ TIL were isolated, as described in *Materials and Methods*. The CD8⁺ TIL samples were mixed and stained with anti-TCR Vβ chains, Vβ2, 4, 7, 10, 13, or 14 mAbs, and subsequently analyzed by flow cytometry. We could not analyze the Vβ1 usage by flow cytometry because no murine anti-Vβ1 Ab was available. Accordingly, we found that CD8⁺/Vβ10⁺ T cells infiltrated at high levels in the tumor (Table I). These results suggested that in the case of P815 tumors, CD8⁺/Vβ10⁺ T cells efficiently infiltrated into the tumor and several identical clones of these T cells had expanded into different areas of the tumor site.
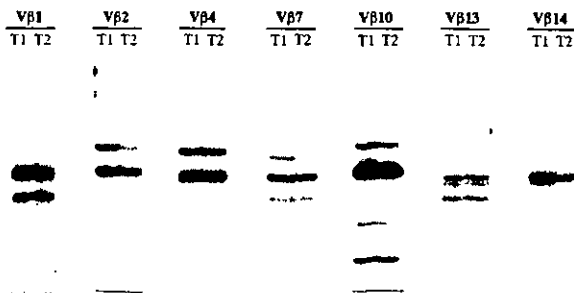

Vβ1	Vβ2	Vβ4	Vβ7	Vβ10	Vβ13	Vβ14
T1 T2	T1 T2	T1 T2	T1 T2	T1 T2	T1 T2	T1 T2

FIGURE 1. Clonality analysis of TCR β-chain messages expressed in P815-infiltrating lymphocytes. Tumor-inoculated samples were obtained from P815-bearing DBA/2 mice and divided into two pieces, called T1 and T2. The RNA was extracted from each sample, and RT-PCR was performed with each of the Vβ-specific primers and a common Cβ primer. The PCR products were analyzed by SSCP. Representative results obtained from selected Vβ amplifications are shown.

Isolation of TCR α- and β-chains derived from a single TIL that expanded into P815 tumors

Because it seemed that CD8⁺/Vβ10⁺ T cells had infiltrated and expanded into the P815 tumors, we attempted to simultaneously isolate TCR α- and β-chains from clonally expanded TIL. To acquire the TCR β-chain massage derived from a single TIL that had expanded into tumors, we performed a single-cell RT-PCR/SSCP study. As shown in Fig. 2A, the resected tumor inoculation samples were divided, and one piece was used for the single-cell sorting. The CD8⁺/Vβ10⁺ TIL were sorted at a ratio of one cell/well. The subsequently synthesized cDNA was amplified by PCR with series sets of Vβ10-specific primers and Cβ common primers. Then we analyzed whether the mobility of the single-cell PCR products on a gel matched that of products from T1 and T2 by SSCP (Fig. 2A). As shown in Fig. 2B, TCR β-chain from clone P09-2C exhibited the same mobility as that from TIL that expanded into both tumor inoculation sites. In addition, to confirm whether the bands represented the identical CDR3 region, we performed sequence analysis. All of the CDR3 regions were identified and found to use the same Vβ10/Dβ1/Jβ2.5 gene segment (Fig. 2C).

Next, we tried to determine the TCR α-chain of the P09-2C clone. The P09-2C cDNA was added poly(A) using TdT, as described previously (24), and PCR was performed with a poly(T) primer and series of Cα common primers. The amplified cDNA was then purified and cloned into a vector, on which we conducted DNA sequencing analysis. Consequently, the TCR α-chain of P09-2C clone was found to consist of the Vα3.5/Jα18 gene segment (Fig. 2C). This finding indicates that CDR3 of the TCR α- and β-chains from clonally expanded TIL could be simultaneously determined by the RT-PCR/SSCP method and the single-cell PCR method.

Reconstituted CD8⁺ T cells have cytotoxic activity against P815 tumor cells

Full-length cDNAs encoding Vα3.5 or Vβ10 TCR that has random CDR3 regions were amplified from cDNA derived from TIL by PCR, and each of the random CDR3 genes was exchanged to the P09-2C CDR3 genes. The full-length P09-2C TCR α- and β-chains were inserted into a retrovirus vector pMX and introduced into splenocytes using recombinant fibronectin fragment CH-296, known as RetroNectin (Fig. 3). Briefly, the supernatants were placed on RetroNectin-coated plates and centrifuged. Using this method, we achieved high efficiency of transduction for T cells. Among CD8⁺ T cells that were introduced, P09-2C TCRαβ-

Table I. *Comparison of Vβ usage from CD8⁺ TIL[a] and spleen cells*

	Vβ TCR	Vβ2	Vβ4	Vβ7	Vβ10	Vβ13	Vβ14
CD8⁺ TIL[b] (%)	93.17 ± 0.98	4.73 ± 3.46	7.97 ± 0.81	2.10 ± 0.53	17.13 ± 5.71	7.40 ± 2.60	2.43 ± 0.87
CD8⁺ spleen cells[b] (%)	95.40 ± 3.40	7.63 ± 0.15	10.17 ± 2.99	3.50 ± 1.06	8.60 ± 0.10	8.90 ± 0.46	5.90 ± 0.61

[a] Tumors were resected from four P815-bearing mice, and a cell suspension was made. The cells were stained with anti-CD8 PE, and CD8⁺ TIL was isolated from the cell suspension. The CD8⁺ TIL sample was mixed and stained with the FITC-conjugated anti-TCR Vβ chain or each of the Vβ mAbs, followed by flow cytometric analysis.
[b] Values represent the means ± SD of nine individual mice.

chain genes, 52.2~75.9% expressed Vβ10 on the cell surface, while mock-introduced T cells were similar to nonintroduced splenocytes (11.4~13.4% of CD8⁺ T cells; data not shown). To test whether the P09-2C TCR-introduced T cells were functional, we measured their cytotoxic effect on P815 tumor cells in a standard ^{51}Cr release assay. As shown in Fig. 4A, P09-2C TCR-introduced T cells positively lysed P815 tumor cells ($\alpha\beta$ vs mock; $p <$ 0.01). In contrast, mock-introduced T cells failed to lyse the tumor cells. Because T cells that express the P09-2C TCR β-chain alone also failed to lyse target tumor cells (Fig. 4B), the cytotoxic ability of the T cells was found to be TCR $\alpha\beta$-chains clonotype specific. These results indicate that CD8⁺ T cells reconstituted with TCR $\alpha\beta$-chains derived from TIL are functional and have a clonotype-specific cytotoxic activity. To determine whether the CTL activity generated by gene transfer is P815 specific, we tested for the CTL activity using a variety of additional target cells. As shown in Fig. 4C, P09-2C TCR-introduced T cells showed no CTL activity against A20 tumor cells (H-2^d), EL-4 (H-2^b), or YAC-1 cells. Furthermore, to determine whether the CTL activity is MHC class I restricted, we performed blocking studies. We found that the anti-

P815 CTL activity was blocked significantly with anti-H2L^d mAb (control IgG vs anti-H2L^d; $p <$ 0.05), but not with anti-H2K^d or anti-H2D^d at a 50:1 E:T ratio (Fig. 4D). These findings suggested that the CTL activity is P815 specific and that this activity is MHC class I restricted.

Reconstituted CD8⁺ T cells accumulate at the P815 tumor site

The P09-2C clone was present and expanded in the P815 tumor site (Fig. 2B), probably due to the Ag recognition by the TCR. Therefore, we examined whether, like the P09-2C clone, the reconstituted T cells accumulated at the P815 tumor site. The pMX-TCRα-IRES/GFP plasmid was used to visualize the TCR α-chains because no murine anti-Vα3 Ab was available. TCR $\alpha\beta$-chains, TCR β-chain plus IRES/GFP, or mock expression vectors were introduced into CD8⁺ T cells. Nude mice were injected s.c. with 1 × 10⁶ tumor cells and i.v. with 2 × 10⁶ infected CD8⁺ T cells. After 12 days, the CD8⁺ TIL were isolated, stained with PE anti-Vβ10, and subsequently analyzed by flow cytometry. The representative result appears in Fig. 5A. Although CD8⁺/GFP⁺/Vβ10⁺ was detected in both the TCR $\alpha\beta$-chains together and the TCR
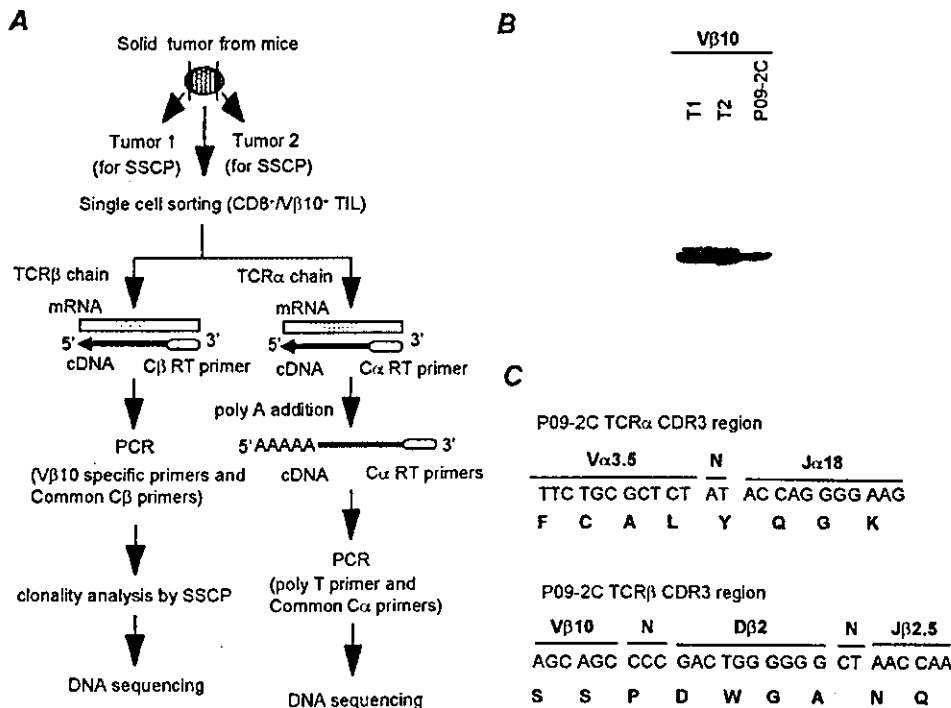

FIGURE 2. Determination of TCR α- and β-chain CDR3 gene segments derived from a CD8⁺ single T cell that clonally expanded in the P815 tumor cells. A, Schematic diagrams of the protocol for the single-cell sorting and the single-cell PCR. B, Analysis of TCR β-chain messages expressed in a single TIL (P09-2C clone). The RNA was extracted from T1, T2, and the P09-2C clone. RT-PCR was then performed with a Vβ10-specific primer and a common Cβ primer for T1 and T2, and with series of the Vβ10-specific primers and common Cβ primers for a P09-2C clone (see *Materials and Methods*). The PCR products were analyzed by SSCP. C, Nucleotide and predicted amino acid sequence from the CDR3 regions of P09-2C clone TCR $\alpha\beta$-chains. The TCRβ bands on an SSCP gel were cut out, amplified, and subsequently sequenced. For DNA sequencing of the P09-2C clone TCR α-chain, the PCR products from a single cell were used. N, the N sequence in each chain.