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

[1]Laboratory for Rheumatic Diseases, SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), 1-7-22, Suehirocho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan. [2]Sankyo, Tokyo, Japan. [3]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. [4]Department of Joint Disease and Rheumatism, Nippon Medical School, Tokyo, Japan. [5]Yukioka Hospital, Osaka, Japan. [6]National Sagamihara Hospital, Kanagawa, Japan. [7]Matsubara Mayflower Hospital, Hyogo, Japan. [8]Osaka Minami National Hospital, Osaka, Japan. [9]Department of Orthopedic Surgery, Tottori University, Tottori, Japan. [10]Yamanashi Prefectural Central Hospital, Yamanashi, Japan. [11]Laboratory for Genotyping and [12]Laboratory for Medical Informatics, SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. [13]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. Correspondence should be addressed to R.Y. (ryamada@src.riken.go.jp).

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*.

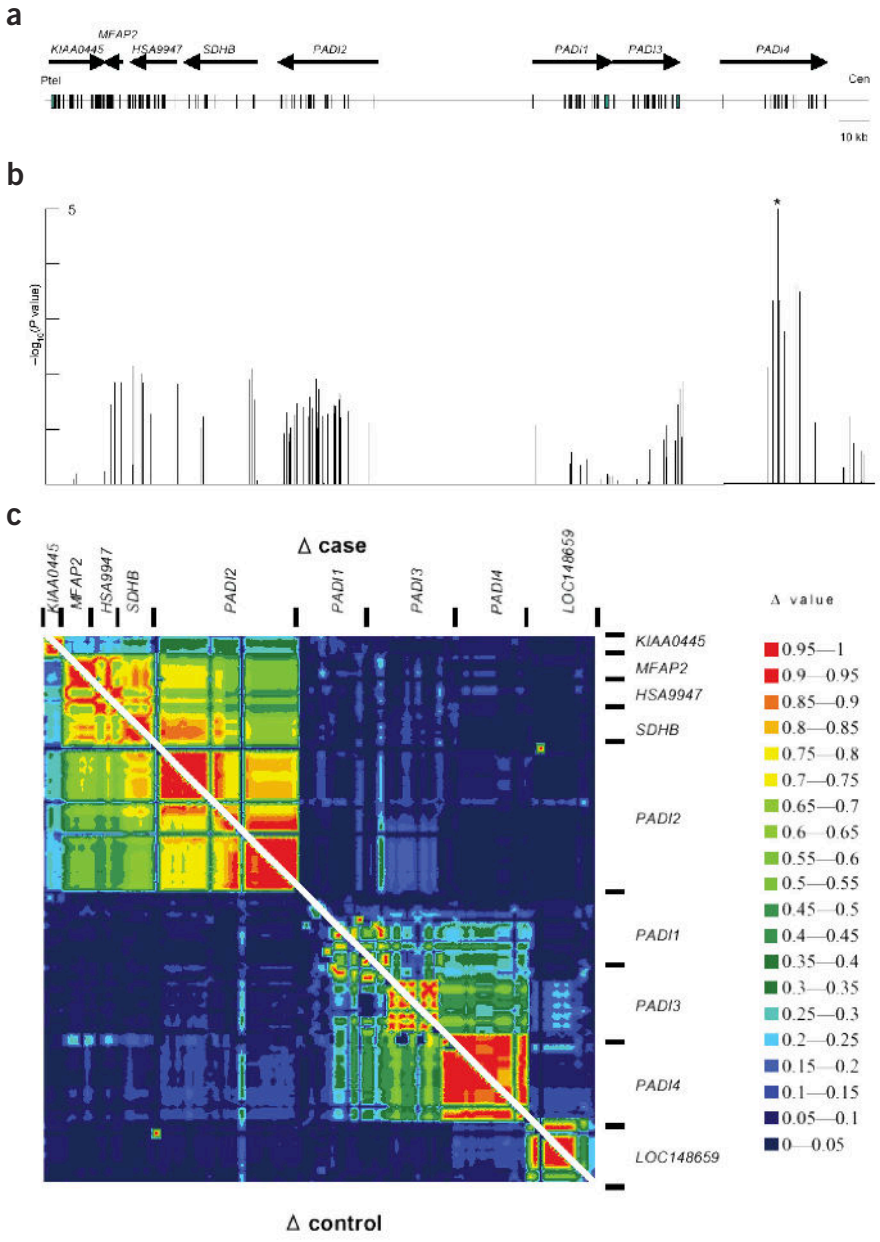

Figure 1 Gene content of NT_034367.1 in chromosome 1p36, case-control association and linkage disequilibrium. (**a**) Genomic structure of genes in this region. Ptel, p telomere; Cen, centromere. (**b**) Case-control association plots ($-\log_{10}(P \text{ value})$) versus location in this region. Asterisk 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.

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 ($\chi^2 = 19.856$, $P = 0.000008$ 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

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

	Genotype of case				Genotype of control				Allele 1 versus allele 2		Genotype 11 versus genotype 22
SNP ID	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**).

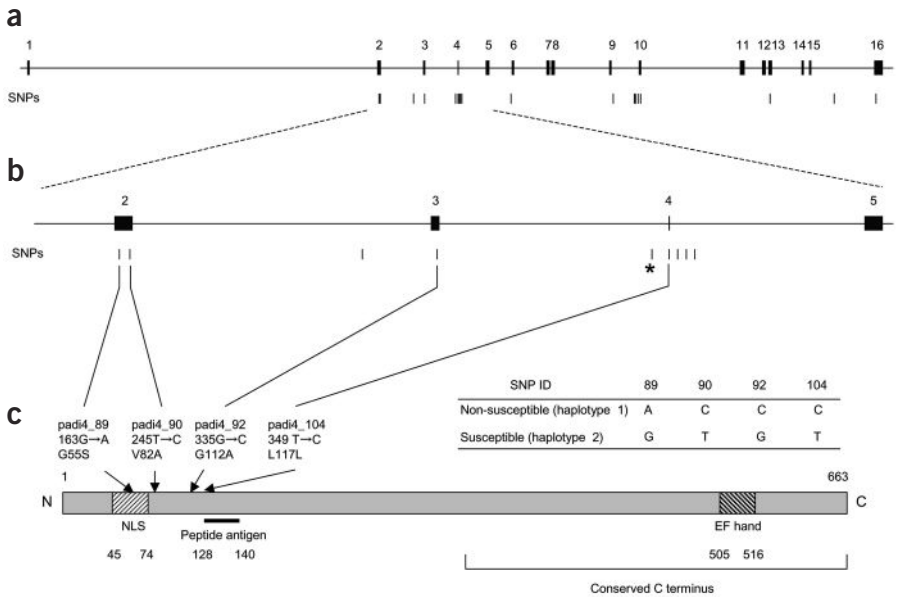

Figure 2 Structure of *PADI4*. (**a**) Exon-intron structure of *PADI4*. SNPs in *PADI4* are indicated below the gene. (**b**) Structure of region including exons 2–5. SNPs in this region are indicated below the gene. The asterisk marks the same SNP that is indicated in **Figure 1b.** (**c**) Protein 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 frequency		SNP ID (as padi4_*x*)																
Haplotype ID	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.

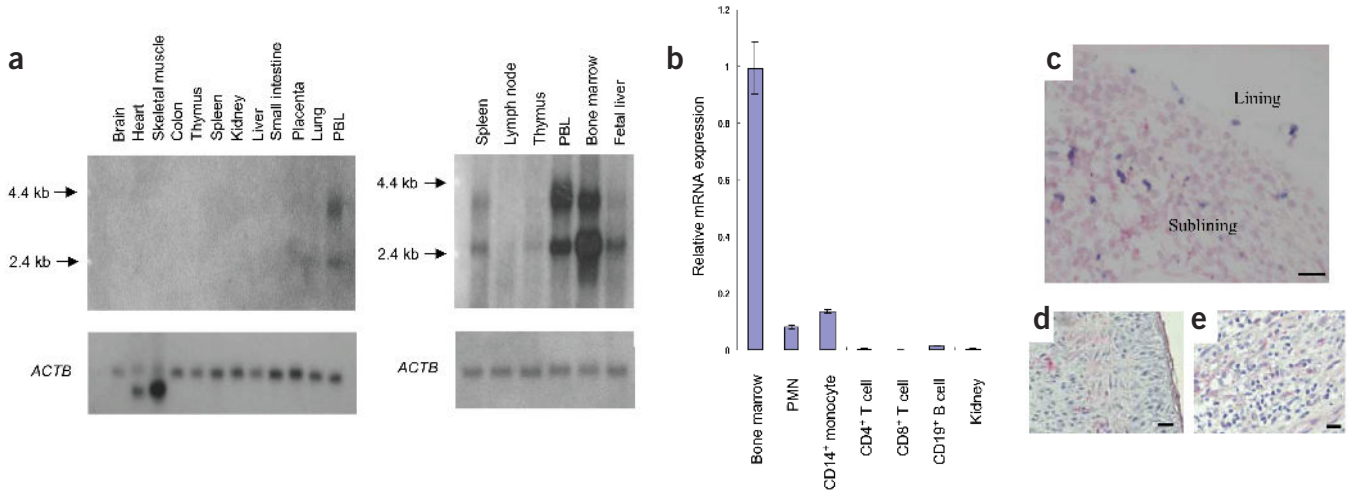

Figure 3 Expression of *PADI4*. (**a**) Expression of *PADI4* mRNA in various normal human tissues. (**b**) Relative expression level of *PADI4* mRNA in normal human tissues and cells. Values represent mean ± s.d. of data from triplicate wells. (**c**) Expression and distribution of *PADI4* mRNA in rheumatoid 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*.

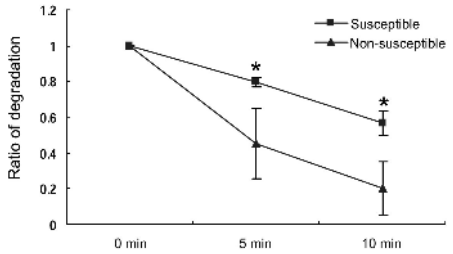

Figure 4 Stability of susceptible and non-susceptible transcripts of *PADI4* mRNA measured as degradation rate. Differences were significant (*$P < 0.05$) after 5 min and 10 min of reaction time. Values represent mean ± s.d. of data from duplicate experiments.

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

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 (BioGate). 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 *Cla*I, 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.

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

ACKNOWLEDGMENTS

We wish to thank E. Tatsu, K. Kobayashi, M. Mito, N. Iwamoto and the other members of the rheumatoid arthritis team for their advice and technical assistance; H. Kawakami for his expertise in computer programming; and many members of the SNP Research Center for helpful discussions and assistance with various aspects of this study. This work was supported by a grant from the Japanese Millennium Project.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 25 February; accepted 2 June 2003
Published online 29 June 2003; doi:10.1038/ng1206

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).

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).