ORIGINAL ARTICLE

Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the *BTNL2*

Shigeki Mitsunaga¹, Kazuyoshi Hosomichi², Yuko Okudaira¹, Hirofumi Nakaoka², Nanae Kunii¹, Yasuo Suzuki³, Masataka Kuwana⁴, Shinji Sato³, Yuko Kaneko⁴, Yasuhiko Homma⁵, Koichi Kashiwase⁶, Fumihiro Azuma⁶, Jerzy K Kulski^{1,7}, Ituro Inoue² and Hidetoshi Inoko¹

The butyrophilin-like protein 2 gene (*BTNL2*) within the class III region of the major histocompatibility complex genomic region was identified as a rheumatoid arthritis (RA) susceptibility gene by exome sequencing (19 RA cases) with stepwise filtering analysis, and then validated by Sanger sequencing and association analysis using 432 cases and 432 controls. Logistic regression of the Sanger-sequenced single-nucleotide variants in an association study of 432 cases and 432 controls showed that 12 non-synonymous single-nucleotide polymorphisms (SNPs) in *BTNL2* were significantly associated with RA. The lowest *P*-values were obtained from three SNPs, rs41521946, rs28362677 and rs28362678, which were in absolute linkage disequilibrium: P = 4.55E - 09, odds ratio = 1.88, 95% confidence interval = 1.52–2.33. The *BTNL2* locates on chromosome 6 between *HLA-DRB1* and *NOTCH4*, and is 170 kb apart from these two genes. Although *DRB1* and *NOTCH4* were reported to be RA-susceptible, the three *BTNL2* SNPs retained significant association with RA when evaluated by the logistic regression with the adjustment for RA-susceptible *HLA-DRB1* alleles in Japanese or rs2071282-T in *NOTCH4*: P = 0.0156 and P = 0.00368, respectively. These results suggest that the three non-synonymous SNPs in *BTNL2* confer RA risk independently from HLA-*DRB1* and *NOTCH4*.

Journal of Human Genetics (2013) 58, 210–215; doi:10.1038/jhg.2013.2; published online 31 January 2013

Keywords: association study; BTNL2; exome sequencing; HLA-DRB1; NOTCH4; rheumatoid arthritis

INTRODUCTION

Rheumatoid arthritis (RA) is a multifactorial and systemic autoimmune disease that can lead to progressive joint destruction and disability. About 60% of the RA risk is genetic¹ and one-third of the genetic risk of RA is attributed to the major histocompatibility complex (MHC, human leukocyte antigens in human: HLA) region² where HLA-DRB1 is strongly associated. Several genome-wide association studies (GWAS) have been conducted based on the 'common-disease common-variant' hypothesis to identify genetic factors that contribute to RA^{3,4} and more than 30 non-HLA loci were reported to be associated with RA.5 However, the identified non-HLA loci so far explain only about 5% of genetic variance.^{5,6} There is a possibility that common variants explain only a modest fraction of the heritability of common diseases, and that rare and low-frequency variants, the so-called 'missing heritability', may explain the remaining genetic variance.^{7,8} The exploration of genetic interactions may also help to overcome some of the problems associated with 'missing heritability?9

Deep resequencing of target regions, which were identified by GWAS, has been performed to identify rare and low-frequency variants. In fact, using PCR products of disease-susceptible loci, many rare variants were identified as disease-susceptible single-nucleotide polymorphisms (SNPs) of type 1 diabetes¹⁰ and inflammatory bowel disease.¹¹ In addition, exome sequencing brought successful results for identification of causative variants of Mendelian diseases.^{12,13} Exomic sequencing with filtering methodology appears to be an efficient strategy for identifying genes involved in Mendelian and/or non-Mendelian complex diseases using a small number of cases.

In the present study, we applied exome sequencing to 19 cases with RA to search for gene-coding variants associated with RA. We identified SNPs within the butyrophilin-like protein 2 gene (*BTNL2*) in the *MHC* region that are likely RA-susceptible gene variants. A functional SNP in *BTNL2* rs2076530 polymorphism was previously associated with RA, but found to have only a secondary effect after taking into account its linkage disequilibrium (LD) with the other

E-mail: hinoko@is.icc.u-tokai.ac.jp

¹Department of Molecular Life Sciences, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Japan; ²Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Mishima, Japan; ³Department of Internal Medicine, Division of Rheumatology, Tokai University School of Medicine, Isehara, Japan; ⁴Department of Internal Medicine, Division of Rheumatology, Keio University School of Medicine, Tokyo, Japan; ⁵Department of Clinical Health Science, Tokai University School of Medicine, Isehara, Japan; ⁶HLA Laboratory, Japanese Red Cross Kanto-Koshinetsu Block Blood Center, Tokyo, Japan and ⁷Centre for Forensic Science, The University of Western Australia, Nedlands, Western Australia

Correspondence: Dr H Inoko, Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan.

Received 10 September 2012; revised 12 December 2012; accepted 14 December 2012; published online 31 January 2013

associated either with a primary effect or as a secondary effect of HLA-DRB1 with many autoimmune diseases, including sarcoidosis,^{15–17} ulcerative colitis,¹⁸ Kawasaki disease,¹⁹ osteoarthritis,^{20,21} systemic lupus erythematosus,¹⁴ multiple sclerosis,²² Graves' disease²³ and Crohn's disease.24

In this paper, we (1) report on finding an association between RA and the BTNL2 gene polymorphisms using exomic and Sanger sequencing in different case and control comparisons after applying different filtering methods, and (2) demonstrate by allele and haplotype frequency analysis in cases and controls that BTNL2 confers a statistically significant RA risk independently from HLA-DRB1 and NOTCH4.

MATERIALS AND METHODS

Subjects

A total of 432 Japanese RA patients were enrolled from outpatients of the Division of Rheumatology, Tokai University Hospital, and Division of Rheumatology, Keio University Hospital. All RA patients fulfilled the 1987 revised criteria of the American College of Rheumatology.²⁵ Two rheumatologists independently evaluated all cases as a blind analysis of clinical information. A total of 432 unrelated healthy Japanese control subjects were recruited from among visitors to the Health Evaluation and Promotion Center of Tokai University Hospital.

All subjects gave written informed consent for genetic screening. Ethical approvals for this study were obtained from the ethics committee of Tokai University School of Medicine and from the ethics committee of Keio University School of Medicine.

DNA samples

DNA samples were extracted from peripheral blood using a DNA extraction kit Genomix (Biologica, Nagoya, Japan) using the manufacturer's instructions.

Next-generation DNA sequencing

Exome sequencing analysis was performed for 19 RA samples, where entire exon sequences were enriched by using a SureSelect Human All Exon kit (Agilent Technologies, Santa Clara, CA, USA), and the libraries were sequenced on the Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) using the pairedend module of 101-bp read. Average read depth was 146, and 92.0, 96.2 and 99.4% of exons were covered on average by at least 20 reads, 10 reads and 1 read, respectively. Sequence reads were aligned to a reference genome (UCSC hg19, NCBI GRCh37) using BWA (Version 0.5.17; http://bio-bwa.sourceforge.net/).26 SAMtools (Version 0.1.18; http://samtools.sourceforge.net/),27 Picard (http://picard.sourceforge.net/) and GATK (http://www.broadinstitute. org/gsa/wiki/index.php/Home_Page)^{28,29} were used for removing duplicated reads, realignment, recalibration and variant identification. Segmental duplication regions were annotated by using ANNOVAR.30

Genotyping of HLA and single-nucleotide variants (SNVs)

We analyzed HLA-DRB1 using the Luminex assay system and HLA typing kits (WAKFlow HLA Typing kits, Wakunaga, Osaka, Japan or LABType SSO, One Lambda, Canoga Park, CA, USA). SNP markers were genotyped by Sanger sequencing using PCR products. The PCR products were purified by ExoSAP-IT treatment (USB, Cleveland, OH, USA) and sequenced using Big Dye Termination chemistry and ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA). The PCR primers used for sequencing and for analyzing SNPs in BTNL2 and NOTCH4 are shown in Supplementary Table 1. We followed NC_000006 (NCBI, GRCh37) for gene structure of BTNL2 (Figure 1).

Statistical analyses

R-software and PLINK (http://pngu.mgh.harvard.edu/purcell/plink/)³¹ were used for statistical analyses: Hardy-Weinberg equilibrium test, Fisher's exact test and logistic regression (additive model). Haplotypes of HLA and SNPs were estimated using PHASE version $2.1.1^{32}$ The D' and the squared Pearson correlation coefficient (r^2) were calculated using the reconstructed haplotypes based on the PHASE analysis. LD blocks were created by the GAB algorithm and the LD patterns were displayed using Haploview software.33

RESULTS

Exomic sequencing analysis

We conducted whole-exome sequencing using 19 RA patients. We excluded synonymous SNVs and SNVs located on the segmental duplication regions. For the quality control of the SNVs, we used only SNVs that fulfilled the following two filtering criteria: (i) the average supporting reads were ≥ 8 ; and (ii) the average SNP quality values generated by the calling algorithm GATK SNV were ≥100. From 57 225 SNV loci, 18 660 SNVs passed the two filtering steps. The summary of the results of the filtered candidate SNVs is shown in Supplementary Table 2 and the flowchart of this study is shown in Supplementary Figure 1.

We postulated that the number of rare and low-frequency variants in a RA-susceptible gene was higher in cases than that in controls.³⁴ Therefore, our first step was to compare the average number of SNVs in each gene between RA patients and 44 disease controls (nonautoimmune diseases). On this basis, we selected the top 30 genes with the highest number of SNV differences between the cases and controls (Supplementary Table 3). Next, we divided the SNVs into two groups: the low variants with an allele frequency of < 0.106 (4 of 38 chromosomes in 19 RA cases) and the common variants with an allele frequency of > 0.106. As a second step, we compared the allele frequency of each SNV in the 19 RA cases with those of 36 Japanese genomes obtained from the data of 1000 genomes project. In this step, we compared the allele frequencies for each SNP in a set of controls different to those used in the first step in order to reduce further the chances of obtaining false-positive associations. This second filtering step narrowed the 30 RA-susceptible candidate genes down to 15 genes with an allele frequency difference of > 0.1 for the common variants and/or higher allele frequency differences for the low variants (Table 1).

The average number of SNV of 99 genes that were reported to be RA-susceptible or were located in the mapped region to be RAsusceptible in 'A Catalog of Published Genome-Wide Association Studies' (http://www.genome.gov/gwastudies/) were shown in Supplementary Table 4. Although only one gene, C6orf10, was appeared in Supplementary Table 3 among 99 genes, the gene was filtered out at the comparison step with 36 Japanese genomes data obtained from the 1000 genomes project (Table 1).

Validation of RA-susceptible gene SNVs by Sanger sequencing and association analysis

The SNVs identified initially by exomic sequencing in the 15 genes listed in Table 1 were validated by Sanger sequencing using 48 cases (excluding the 19 exome-sequenced samples) and 48 controls (Table 1). Then the differences in the allele frequencies between the cases and controls were determined by the Fisher's exact test. Only the SNVs located in BTNL2, NOTCH4 and MYPN showed *P*-value < 0.05.

We further confirmed the association of these SNVs with RA by using 432 cases, including the 19 exome-sequenced samples, and 432 controls. Among these samples, 411 RA patients and 420 control





Figure 1 Schematic illustration of the location of the BTNL2 and the non-synonymous SNPs in BTNL2 observed in this study.

Table 1 Top 15 gene SNVs and their overall significance after Sanger
sequencing and association analysis between 48 RA cases and 48
controls

Gene (symbol)	Cytogenetic location	Difference in average number of SNV ^a	Number of SNV ^b	Significance after Sanger sequencing ^c
BTNL2	6p21.32	6.3	16	P=0.0177
CDC27	17q21.32	6	14	Not SNP
PRRC2A	6p21.33	3.7	9	P>0.1
CTBP2	10q26.13	3	13	Not SNP
ZNF83	19q13.41	2.8	13	P>0.1
MUC21	6p21.32	2.5	19	Repeat sequences
VARS2	6p21.33	2	9	P>0.1
CCHCR1	6p21.33	1.9	14	P>0.1
UBD	6p22.1	1.8	6	P>0.1
NOTCH4	6p21.32	1.5	10	P=0.0116
OR10C1	6p22.1	1.4	7	P>0.1
PRSS2	7q35	1.3	7	Not SNP
KIAA1683	19p13.1	1.3	10	P>0.1
EFCAB4B	12p13.32	1.1	5	Not SNP
MYPN	10q21.3	1	6	P=0.0272

Table 2	Association	analysis with	rheumatoid	arthritis	using logistic
regressio	on				
					· · · · · · · · · · · · · · · · · · ·

Gene	Exon/intron	SNP/allele	allele Base OR 95% Cl		P-value	
NOTCH4	Intron 4	rs2071284	А	3.14	2.27–4.33	4.02E-12
	Exon 4	rs2071283	А	3.14	2.27-4.33	4.02E-12
		rs2071282	Т	3.14	2.27-4.33	4.02E-12
		rs2071281	Т	3.14	2.27-4.33	4.02E-12
		rs415929	G	1.27	1.05-1.56	2.17E-02
BTNL2	Exon 6	rs28362675	Α	1.93	1.54–2.42	1.58E-08
		rs41521946	Т	1.88	1.52-2.33	4.55E-09
		rs28362677	Т	1.88	1.52-2.33	4.55E-09
		rs28362678	Α	1.88	1.52-2.33	4.55E-09
	Exon 5	rs2076530	G	1.47	1.21-1.78	1.09E-04
		rs41441651	А	1.93	1.53-2.44	3.98E –08
		rs41417449	G	1.92	1.53-2.44	4.40E-08
		rs34423804	Т	1.92	1.53-2.44	4.40E-08
		rs41355746	А	1.93	1.53-2.44	3.98E –08
	Exon 3	rs28362680	Т	1.46	1.20-1.78	2.06E04
		rs2076523	А	0.72	0.59–0.87	8.54E-04
		rs78587369	Т	1.94	1.54-2.45	2.62E-08
DRB1		DRB1*04:05	А	2.92	2.22–3.84	1.70E-14

Abbreviations: SNP, single-nucleotide polymorphism; SNV, single-nucleotide variant. ^aThe difference in the average number of SNVs in each gene between 19 cases and 44 controls.

^bThe number of SNVs in cases.

 $^{\rm c}{\rm The}$ P-value indicates the first screening result using 48 cases and 48 controls.

samples with complete genotypes for these SNPs were used for the analyses. Table 2 shows the *BTNL2* and *NOTCH4* SNPs that were significantly associated with RA, although rs2071282 in *NOTCH4* was already reported to be RA-susceptible.³⁵ The rs10997975 in *MYPN* showed weak association: P = 0.0047, odds ratio = 1.33, 95% confidence interval = 1.09–1.62 by the Fisher's exact test. As the association of rs10997975 in *MYPN* with RA was weak and *BTNL2* and *NOTCH4* are located in close proximity to *HLA-DRB1*, we then focused mainly on *BTNL2* as well as analyzing the relationships between *BTNL2*, *NOTCH4* and *HLA-DRB1*.

Twelve non-synonymous SNPs in *BTNL2* were significantly associated with RA (Table 2) by logistic regression. The three SNPs, rs28362678, rs28362677 and rs41521946, in exon 6 were in absolute LD ($r^2 = 1$, see below) and showed the lowest *P*-value among the 12 SNPs.

Abbreviations: CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism. All of the SNPs listed here except for four SNPs, rs2071284, rs2071283, rs2071281 and rs415929, are non-synonymous.

As the *BTNL2* rs2076530 in exon 5 was previously reported to be a RA-susceptible SNP that was insignificant when tested conditionally on *HLA DQB1–DRB1* haplotypes,¹⁴ we conducted logistic regression with the adjustment for RA-susceptible *DRB1* alleles in Japanese (*DRB1*04:05, DRB1*04:01* and *DRB1*10:01*)³⁶ (Table 3) or the RA-susceptible *NOTCH4* allele (Table 4). The three SNPs, rs28362678, rs28362677 and rs41521946, of exon 6 retained significant *P*-values (*P*<0.05) in both of logistic regression with the adjustment, although the other SNPs, including rs2076530, in exon 5 were no longer significantly associated with RA (*P*>0.05).

The amino acids substituted by the three non-synonymous SNPs were shared by other species, including orangutan (Supplementary Figure 2), that is, the substituted amino acids that were found in the human were conserved in other mammalian species. It implies the

Gene	Exon/intron	SNP/allele	Base	OR	95% CI	P-value
NOTCH4	Intron 4	rs2071284	А	1.39	0.89–2.17	0.146
	Exon 4	rs2071283	А	1.39	0.89–2.17	0.146
		rs2071282	Т	1.39	0.89–2.17	0.146
		rs2071281	Т	1.39	0.89–2.17	0.146
		rs415929	G	0.94	0.75-1.17	0.569
BTNL2	Exon 6	rs28362675	Α	1.13	0.86-1.50	0.382
		rs41521946	Т	1.34	1.06-1.70	1.45E02
		rs28362677	Т	1.34	1.06-1.70	1.45E02
		rs28362678	А	1.34	1.06-1.70	1.45E02
	Exon 5	rs2076530	G	0.92	0.74-1.15	0.465
		rs41441651	А	1.08	0.81-1.44	0.615
		rs41417449	G	1.07	0.80-1.43	0.644
		rs34423804	Т	1.07	0.80-1.43	0.644
		rs41355746	А	1.08	0.81-1.44	0.615
	Exon 3	rs28362680	Т	1.06	0.85-1.33	0.593
		rs2076523	А	1.04	0.83-1.29	0.759
		rs78587369	Т	1.07	0.80-1.44	0.650
DRB1		DRB1*04:05	А	NA	NA	NA

Table 3 Logistic regression with the adjustment for RA-risk HLA-DRB1 alleles^a

Table 4 Logistic regression with the adjustment for rs2071281 in exon 4 of *NOTCH4*^a

Gene	Exon/intron	SNP/allele	Base	OR	95% CI	P-value
NOTCH4	Intron 4	rs2071284	А	NA	NA	NA
	Exon 4	rs2071283	А	NA	NA	NA
		rs2071282	Т	NA	NA	NA
		rs2071281	Т	NA	NA	NA
		rs415929	G	0.83	0.65–1.05	0.124
BTNL2	Exon 6	rs28362675	А	1.31	0.99–1.72	0.055
		rs41521946	Т	1.42	1.12-1.80	3.68E –03
		rs28362677	Т	1.42	1.12-1.80	3.68E –03
		rs28362678	А	1.42	1.12-1.80	3.68E-03
	Exon 5	rs2076530	G	1.18	0.96-1.45	0.118
		rs41441651	А	1.26	0.94-1.68	0.119
		rs41417449	G	1.25	0.94-1.67	0.121
		rs34423804	G	1.25	0.94-1.67	0.121
		rs41355746	А	1.26	0.94-1.68	0.119
	Exon 3	rs28362680	Т	1.10	0.88-1.37	0.401
		rs2076523	Α	0.92	0.75-1.14	0.451
		rs78587369	Т	1.27	0.95–1.69	0.104
DRB1		DRB1*04:05	А	2.16	1.46–3.21	1.12E-04

Abbreviations: CI, confidence interval; OR, odds ratio; NA, not applicable; SNP, singlenucleotide polymorphism.

^aAdjustment for DRB1*04:05, DRB1*04:01 and DRB1*10:01.

function of the BTNL2 protein was not changed or affected drastically by evolution and perhaps the selection pressure was weak.

LD and haplotype analysis between BTNL2, NOTCH4 and HLA-DRB1

As only the three SNPs in exon 6 of BTNL2 retain significant association with RA in logistic regression with the adjustment for RAsusceptible DRB1 alleles or NOTCH4, we compared LD values among the BTNL2 and NOTCH4 SNPs and DRB1*04:05 using phased haplotypes (Table 5 and Supplementary Figure 3). Both values of D' and r^2 between rs2076530 and DRB1*04:05 showed lower values compared with those between the three BTNL2 SNPs and DRB1*04:05: D' = 0.718, $r^2 = 0.124$ for rs2076530—DRB1*04:05, and D' = 0.807, $r^2 = 0.268$ for rs28362678/rs28362677/rs41521946-DRB1*04:05, respectively. On the other hand, the MAF (minor allele frequency) and allele frequency were 0.500 for rs2076530, 0.369 for the 3 BTNL2 SNPs and 0.194 for DRB1*04:05, respectively. Therefore, the differences in MAF may explain the lower D' and r^2 obtained between rs2076530 and DRB1*04:05 compared with those between the three BTNL2 SNPs and DRB1*04:05. That is, only a minority of the haplotypes carrying rs2076530-G have DRB1*04:05.

The comparison of the number of *BTNL2* SNVs in the patients and controls with respect to *HLA-DRB1* genotypes was also shown in the Supplementary Table 5.

DISCUSSION

We showed here that the three non-synonymous SNPs, rs28362678, rs28362677 and rs41521946, of the exon 6 of *BTNL2*, which were in absolute LD with each other, were associated with RA independently from *HLA-DRB1* and *NOTCH4*. Although the other nine SNPs, including rs2076530, of *BTNL2* also showed significant association with RA, they were no longer significant in the logistic regression with the adjustment for RA-susceptible *HLA-DRB1* alleles in Japanese:

Abbreviations: CI, confidence interval; OR, odds ratio; NA, not applicable; SNP, singlenucleotide polymorphism.

The SNPs rs2071281, rs2071282, rs2071283 and rs2071284 are in complete linkage disequilibrium.

DRB1*04:05, DRB1*04:01 and DRB1*10:01. These results agree with the report that the association between rs2076530 and RA was attributed to LD with DR-DQ haplotypes.¹⁴ However, this is the first report on the association between non-synonymous SNPs other than rs2076530 in *BTNL2* with RA. Although the association of rs2076530 with RA was attributed to the LD with *DRB1*, the LD values, D' and r^2 , of rs2076530 were lower than those of the three *BTNL2* SNPs in exon 6. The differences in MAF appear to explain the phenomenon.

In this study, the association of *NOTCH4* with RA was a secondary effect to that of *HLA-DRB1* based on the logistic regression with the adjustment for *DRB1*04:05*, although it was reported previously that the association of *NOTCH4* with RA was independent from *HLA-DRB1* using a partially recessive model focusing on a shared epitope of *HLA-DRB1.*³⁵ We also showed here a weak but novel association between *MYPN* and RA. The *MYPN* gene, located on chromosome 10, encodes the myopallidin protein that interacts with Actinin, alpha 2, ANKRD23 and ANKRD1 and is a component of a family of titin filament-based stress/strain response molecules in myofibrils.^{37,38} It is difficult to envisage what role the myopallidin protein, as a component of the sarcomere, might have in RA. Therefore, the association of MYPN with RA should be retested and confirmed in a replication study using many more samples.

To date, a number of GWAS were conducted on RA. However, there were few reports on the association between *BTNL2* and RA. The r^2 between SNPs of *BTNL2* ranged between 0.084 and 1, and the lowest r^2 (0.084) was between rs2076530 and rs28362680 (Table 5). It is likely that SNPs around *BTNL2* on the GWAS platforms do not fully represent the variants that can be found within the *BTNL2*-LD block and that were detected in this study. For example, the two probes corresponding to rs3817969 and rs1980493, which are 1.1 kb upstream of *BTNL2* and in intron 5 of *BTNL2*, respectively, are used on the OmniExpress (Illimina) across the exon 6 of *BTNL2*. The SNP

Table 5 LD values between SNPs in the NOTCH—HLA-DRB1*04:05 region

	BTNL2								0001			
	NOTCH4		NOTCH4		Exc	on 6	Ex	on 5		Exon 3		DKBI
	rs2071281	rs415929	rs28362675	rs41521946	rs2076530	rs41441651	rs28362680	rs2076523	rs78587369	*04:05		
rs2071281		0.991	0.934	0.952	0.843	0.943	0.933	0.927	0.936	0.952		
rs415929	0.265		0.241	0.322	0.388	0.254	0.070	0.198	0.253	0.555		
rs28362675	0.352	0.039		0.996	0.908	0.971	0.933	0.923	0.971	0.797		
rs41521946	0.239	0.101	0.648		0.915	0.996	0.671	0.613	1	0.807		
rs2076530	0.110	0.086	0.315	0.490		0.932	0.340	0.500	0.895	0.718		
rs41441651	0.384	0.040	0.881	0.605	0.310		0.981	0.975	0.981	0.793		
rs28362680	0.185	0.002	0.457	0.363	0.084	0.473		0.964	0.986	0.751		
rs2076523	0.127	0.023	0.312	0.211	0.240	0.325	0.648		0.981	0.790		
rs78587369	0.379	0.040	0.881	0.610	0.286	0.963	0.478	0.330		0.803		
DRB1*04:05	0.582	0.130	0.400	0.268	0.124	0.424	0.187	0.144	0.435			

Abbreviations: LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

The three SNPs, rs2071282, rs2071283 and rs2071284, the two SNPs, rs28362677 and rs28362678, and the three SNPs, rs41355746, rs34423804 and rs41417449, which were in complete linkage disequilibrium ($r^2 = 1$) with rs2071281, with rs41521946 and with rs41441651, respectively, in cases were represented by the first one of each complete linkage group. The upper-right half indicates D' values and the lower-left half indicates r^2 .

rs28362675, which showed no significant association with RA in the logistic regression with the adjustment for $DRB1^*04:05$, is more closely positioned to rs3817969 than to the three BTNL2 SNPs of exon 6 that were in absolute LD. In addition, the MAF of rs1980493 is 0.085 in Japanese (HapMap data) and is considerably lower than that of the three BTNL2 SNPs (MAF: 0.369). Therefore, the tagging of only two SNPs for BTNL2 in the OmniExpress arrays may not work well for detecting the three BTNL2 SNPs in exon 6. In this regard, exome sequencing will complement the tagging SNPs in the OmniExpress arrays that might not be representative of the BTNL2 LD blocks shown here.

BTNL2 is a member of the immunoglobulin gene superfamily with homology to the B7 costimulatory molecule and consists of two sets of two Ig domains, IgV and IgC.³⁹ BTNL2 mRNA is highly expressed in lymphoid tissues as well as in the intestine and is involved in the regulation of T-cell activation as a negative costimulatory molecule.^{39,40} Interestingly, the CTLA4 gene, which encodes for a protein, which is one of the ligands of B7 and a inhibitory signal to activated T cells, is also associated with RA.41,42 Therefore, it is likely that the variants in BTNL2 confer risk to autoimmune disease, although the amino acids substituted by the three SNPs would not be expected to result in drastic changes because the changed amino acids are shared by other species, including orangutan (Supplementary Figure 1). In fact, there are many reports on the associations of BTNL2 SNPs with autoimmune disease, sarcoidosis,15-17 ulcerative colitis,¹⁸ Kawasaki disease¹⁹ and osteoarthritis.^{20,21} The two intergenic SNPs located between BTNL2 and DRA were reported to be associated with anti-cyclic citrullinated peptide antibody titer in adults with RA.⁴³ In these diseases, the independence from HLA-DRB1 is often controversial. However, the majority of these associations were described for the SNP rs2076530, intron SNPs and intergenic SNPs. However, other BTNL2 non-synonymous SNPs were associated with sarcoidosis15 and ulcerative colitis.18 In sarcoidosis, the association of the three non-synonymous SNPs in exon 6 were secondary effects based on LD with rs2076530. In ulcerative colitis, the association was with rs2076523 in exon 3. In our study on RA, however, the P-values of the three SNPs were lower than for both rs2076530 and rs2076523 (Table 2). In addition, the associations of rs2076530 and rs2076523 were owing to LD with DRB1 (Table 3). The association studies using

three non-synonymous SNPs in exon 6 of *BTNL2* on auto-immune disease as well as replication studies for RA may clarify the association of *BTNL2* with auto-immune diseases, including systemic lupus erythematosus,¹⁴ multiple sclerosis,²² Graves' disease²³ and Crohn's disease,²⁴ for which the association with rs2076530 was reported to be a secondary effect to *DRB1*.

In this study on the identification of RA-susceptible genes, we performed exome sequencing followed by the validation of the candidate disease SNVs in case-control association analysis. In order to identify the candidate genes, we compared the average number of SNV per gene based on the hypothesis that the number of rare and low-frequency variants in a disease-susceptible gene in cases would be higher than that in controls. More than half of genes listed in Supplementary Tables 3 and 1 were located on 6p21-22, in which region the strongest RA-susceptible gene, HLA-DRB1, was also located. We could not deny the effects of the LD with HLA-DRB1 on the filtering process. Therefore, the testing of independence from HLA-DRB1 and adjacent RA-susceptible genes was required, and we carried out such testing. Eventually, we could not obtain RAsusceptible rare and low-frequency variants in this study. It is likely that the presence of common variants in the same gene interfered with the detection of rare and low-frequency variants, and that a limited number of exome-sequenced cases affected the efficiency of detecting rare and low-frequency variants. We expect that a greater accumulation of exome-sequenced cases will resolve these limitations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the DNA donors and supporting medical staff for making this study possible. We thank Miyako Nakagawa and Tomomi Ito (Division of Rheumatology, Tokai University School of Medicine), Hisako Kawata, Masayuki Tanaka and Hideki Hayashi (Education and Research Support Center, Tokai University School of Medicine) for their excellent technical assistance. We also thank Dr Atsushi Toyoda (Comparative Genomics Laboratory, National Institute of Genetics), Dr Shoji Tsuji, Dr Jun Mitsui and Dr Hiroyuki Ishiura (Department of Neurology, Division of Neuroscience, Graduate School of Medicine, The University of Tokyo), Dr Shinichi Morishita

215

and Dr Jun Yoshimura (Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo) for their supporting based on the Grant in-Aid-for Scientific Research on Innovative Areas 'Genome Science' (No.221S0002) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was supported in part by a Grant-in-Aid for Scientific Research (C) (No. 22510213) and a Grant-in-Aid for Scientific Research from the Japanese Ministry of Health, Labor, and Welfare.

- MacGregor, A. J., Snieder, H., Rigby, A. S., Koskenvuo, M., Kaprio, J., Aho, K. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
- 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 Wellcome Trust Case Control Consortium. Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls. *Nature* 447, 661–678 (2007).
- 4 Wellcome Trust Case Control ConsortiumCraddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V. *et al.* Genome-wide association study of CNVs in 16000 cases of eight common diseases and 3000 shared controls. *Nature* **464**, 713–720 (2010).
- 5 de Vries, R. Genetics of rheumatoid arthritis: time for a change! *Curr. Opin. Rheumatol.* 23, 227–232 (2011).
- 6 Raychaudhuri, S., Remmers, E. F., Lee, A. T., Hackett, R., Guiducci, C., Burtt, N. P. et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat. Genet. 40, 1216–1223 (2008).
- 7 Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J. et al. Finding the missing heritability of complex diseases. *Nature* 461, 747–753 (2009).
- 8 McClellan, J. & King, M. C. Genetic heterogeneity in human disease. Cell 141, 210–217 (2010).
- 9 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* 109, 1193–1198 (2012).
- 10 Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324, 387–389 (2009).
- 11 Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43**, 1066–1073 (2011).
- 12 Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C. et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276 (2009).
- 13 Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M. et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 42, 30–35 (2010).
- 14 Orozco, G., Eerligh, P., Sánchez, E., Zhernakova, S., Roep, B. O., González-Gay, M. A. et al. Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. *Hum. Immunol.* 66, 1235–1241 (2005).
- 15 Valentonyte, R., Hampe, J., Huse, K., Rosenstiel, P., Albrecht, M., Stenzel, A. *et al.* Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat. Genet.* **37**, 357–364 (2005).
- 16 Rybicki, B. A., Walewski, J. L., Maliarik, M. J., Kian, H. & Iannuzzi, M. C. ACCESS Research Group. The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. Am. J. Hum. Genet. 77, 491–499 (2005).
- 17 Wijnen, P. A., Voorter, C. E., Nelemans, P. J., Verschakelen, J. A., Bekers, O. & Drent, M. Butyrophilin-like 2 in pulmonary sarcoidosis: a factor for susceptibility and progression? *Hum. Immunol.* **72**, 342–347 (2011).
- 18 Pathan, S., Gowdy, R. E., Cooney, R., Beckly, J. B., Hancock, L., Guo, C. *et al.* Confirmation of the novel association at the BTNL2 locus with ulcerative colitis. *Tissue Antigen* **74**, 322–329 (2009).
- 19 Hsueh, K. C., Lin, Y. J., Chang, J. S., Wan, L. & Tsai, F. J. BTNL2 gene polymorphisms may be associated with susceptibility to Kawasaki disease and formation of coronary artery lesions in Taiwanese children. *Eur. J. Pediatr.* **169**, 713–719 (2010).
- 20 Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J. J., Furuichi, T. *et al.* New sequence variants in HLA class II/III region associated with

susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One* **5**, e9723 (2010).

- 21 Valdes, A. M., Styrkarsdottir, U., Doherty, M., Morris, D. L., Mangino, M., Tamm, A. et al. Large scale replication study of the association between HLA class II/BTNL2 variants and osteoarthritis of the knee in European-descent populations. *PLoS One* 6, e23371 (2011).
- 22 Traherne, J. A., Barcellos, L. F., Sawcer, S. J., Compston, A., Ramsay, P. P., Hauser, S. L. *et al.* Association of the truncating splice site mutation in BTNL2 with multiple sclerosis is secondary to HLA-DRB1*15. *Hum. Mol. Genet.* **15**, 155–161 (2006).
- 23 Simmonds, M. J., Heward, J. M., Barrett, J. C., Franklyn, J. A. & Gough, S. C. Association of the BTNL2 rs2076530 single nucleotide polymorphism with Graves' disease appears to be secondary to DRB1 exon 2 position beta74. *Clin. Endocrinol.* 65, 429–432 (2006).
- 24 Johnson, C. M., Traherne, J. A., Jamieson, S. E., Tremelling, M., Bingham, S., Parkes, M. *et al.* Analysis of the BTNL2 truncating splice site mutation in tuberculosis, leprosy and Crohn's disease. *Tissue Antigen* **69**, 236–241 (2007).
- 25 Arnett, F. C., Edworthy, S. M., Bloch, D. A., McShane, D. J., Fries, J. F., Cooper, N. S. et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. **31**, 315–324 (1988).
- 26 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 754–760 (2009).
- 27 Li, H., Handsaker, B., Wysoker., A., Fennell, T., Ruan, J., Homer, N. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- 28 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome. Res.* **20**, 1297–1303 (2010).
- 29 DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- 30 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010).
- 31 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 32 Stephens, M. & Donnelly, P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
- 33 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265 (2005).
- 34 Bansal, V., Libiger, O., Torkamani, A. & Schork, N. J. Statistical analysis strategies for association studies involving rare variants. *Nat. Rev. Genet.* 11, 773–785 (2010).
- 35 Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K. et al. Whole genome association study of rheumatoid arthritis using 27039 microsatellites. *Hum. Mol. Genet.* 14, 2305–2321 (2005).
- 36 Mitsunaga, S., Suzuki, Y., Kuwana, M., Sato, S., Kaneko, Y., Homma, Y. et al. Associations between six classical HLA loci and rheumatoid arthritis: a comprehensive analysis. *Tissue Antigen* 80, 16–25 (2012).
- 37 Bang, M. L., Mudry, R. E., McElhinny, A. S., Trombitás, K., Geach, A. J., Yamasaki, R. et al. Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. J. Cell. Biol. 153, 413–427 (2001).
- 38 Miller, M. K., Bang, M. L., Witt, C. C., Labeit, D., Trombitas, C., Watanabe, K. *et al.* The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules. *J. Mol. Biol.* **333**, 951–964 (2003).
- 39 Arnett, H. A., Escobar, S. S., Gonzalez-Suarez, E., Budelsky, A. L., Steffen, L. A., Boiani, N. *et al.* BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. *J. Immunol.* **178**, 1523–1533 (2007).
- 40 Nguyen, T., Liu, X. K., Zhang, Y. & Dong, C. BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. J. Immunol. 176, 7354–7360 (2006).
- 41 Gregersen, P. K., Amos, C. I., Lee, A. T., Lu, Y., Remmers, E. F., Kastner, D. L. *et al.* REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
- 42 Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
- 43 Cui, J., Taylor, K. E., Destefano, A. L., Criswell, L. A., Izmailova, E. S., Parker, A. et al. Genome-wide association study of determinants of anti-cyclic citrullinated peptide antibody titer in adults with rheumatoid arthritis. *Mol. Med.* **15**, 136–143 (2009).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)