Department of Medicine in Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden. [30]Arthritis Research Campaign–Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [31]Division of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada. [32]Commisariat a l'Energie Atomique (CEA), Institut Genomique, Centre National de Genotypage, Evry, France. [33]Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain, Paris, France. [34]Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan. [35]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [36]Laboratory for International Alliance, CGM, RIKEN, Yokohama, Japan. [37]Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto, Japan. [38]Core Research for Evolutional Science and Technology (CREST) Program, Japan Science and Technology Agency, Kawaguchi, Japan. [39]Institut National de la Santé et de la Recherche Médicale (INSERM), Unité U852, Kyoto University Graduate School of Medicine, Kyoto, Japan. [40]These authors contributed equally to this work. [41]These authors jointly directed this work. Correspondence should be addressed to Y.K. (ykochi@src.riken.jp) or K.O. (ohmurako@kuhp.kyoto-u.ac.jp).

196

ONLINE METHODS

Subjects. The Japanese participants in the meta-analysis (4,074 rheumatoid arthritis cases and 16,891 controls) and the replication study (5,277 rheumatoid arthritis cases and 21,684 controls) were obtained through the collaborations of the GARNET consortium (**Supplementary Table 1**)[10,12]. The meta-analysis was conducted on three independent GWAS (from the BioBank Japan Project[18] with 2,414 rheumatoid arthritis cases and 14,245 controls[10], Kyoto University with 1,237 rheumatoid arthritis cases and 2,087 controls[12] and IORRA[19] with 423 rheumatoid arthritis cases and 559 controls). The replication study consisted of two independent cohorts (cohort 1 included 3,830 rheumatoid arthritis cases and 17,920 controls, and cohort 2 included 1,447 rheumatoid arthritis cases and 3,764 controls). We employed a case-control cohort of SLE (891 cases and 3,384 controls)[22] and 1,783 cases with Graves' disease[10]. Details of 5,539 rheumatoid arthritis cases and 20,169 controls included in the meta-analysis in European populations were described elsewhere[15]. All participants provided written informed consent for participation in the study, as approved by the ethical committees of the institutional review boards. Detailed descriptions of the participating subjects are provided (**Supplementary Note**).

Genotyping and quality control in the GWAS. Genotyping platforms and quality control criteria for the GWAS, including cutoff values for sample call rates, SNP call rates, MAF and Hardy-Weinberg *P* values, are given (**Supplementary Table 2**). For the subjects enrolled in each of three GWAS, we excluded closely related subjects with first- or second-degree kinship, which was estimated using PLINK version 1.06 (see URLs). We also excluded the subjects determined to be ancestry outliers from East Asian populations using PCA performed by EIGENSTRAT version 2.0 (see URLs) along with HapMap Phase 2 panels (release 24; **Supplementary Fig. 1**). Genotype imputation was performed on the basis of the HapMap Phase 2 East Asian populations, using MACH version 1.0.16 (see URLs) in a two-step procedure as described elsewhere[25]. We excluded imputed SNPs with MAF < 0.01 or *Rsq* < 0.5 from each of the GWAS. Associations of the SNPs with rheumatoid arthritis were assessed by logistic regression models assuming additive effects of the allele dosages of the SNPs using mach2dat software (see URLs).

Meta-analysis. We included 1,948,139 autosomal SNPs that satisfied quality control criteria in all three GWAS (**Supplementary Table 2**). SNP information was based on a forward strand of the NCBI build 36.3 reference sequence. The meta-analysis was performed using an inverse variance method assuming a fixed-effects model from the study-specific effect sizes (logarithm of odds ratio) and the standard errors of the coded alleles of the SNPs determined with the Java source code implemented by the authors[25]. Genomic control corrections[26] were carried out on test statistics of the GWAS using the study-specific inflation factor (λ_{GC}) and was applied or reapplied to the results of our current meta-analysis (**Supplementary Fig. 2**).

Replication study. We selected a SNP for the replication study from each of the loci that exhibited $P < 5.0 \times 10^{-4}$ in the meta-analysis that had not previously been reported as rheumatoid arthritis susceptibility loci[1–16] (**Supplementary Table 3**). For control subjects, we used genotype data obtained from additional GWAS for non-autoimmune diseases or healthy controls, genotyped using Illumina HumanHap550 BeadChips or HumanHap610-Quad BeadChips, and

the cases for rheumatoid arthritis and Graves' disease were genotyped with the TaqMan genotyping system (Applied Biosystems; **Supplementary Table 1**). Selection of the SNP was conducted according to the following criteria: if the SNP with the most significant association in the locus was genotyped in the replication control panel, then that SNP was selected; otherwise, a tag SNP in the replication control panel with the strongest LD was selected (mean $r^2 = 0.89$). For the three SNPs that yielded low call rates (<90%), we alternatively selected proxy SNPs with the second strongest LD. As a result, average genotyping call rates of the SNPs were 99.9% and 99.0% for the controls and cases, respectively. We then evaluated concordance rates between the assayed genotypes by applying these two different methods to samples from 376 subjects who were randomly selected. This procedure yielded high concordance rates of ≥99.9%. Associations of the SNPs were evaluated using logistic regression assuming an additive-effects model of genotypes in R statistical software version 2.11.0 (see URLs). The combined study of the meta-analysis and replication study was performed using an inverse variance method assuming a fixed-effects model[25].

***Cis* eQTL analysis.** For each marker SNP of the newly identified rheumatoid arthritis susceptibility locus, correlations between SNP genotypes and expression levels of genes located 300 kb upstream or downstream of the SNP measured in B-lymphoblastoid cell lines (GSE6536) were evaluated using data from the HapMap Phase 2 east Asian populations[27].

Multi-ancestry analysis of the meta-analyses in Japanese and Europeans. We evaluated the associations of the variants in the validated rheumatoid arthritis susceptibility loci by comparing the results from the current meta-analysis in Japanese with those from a previous meta-analysis in Europeans[15]. We assessed two variants in the *IRF5* locus, where different causal variants were identified in the two populations[24]. For the conditional analysis of the regional associations in the *ARID5B* locus (**Supplementary Fig. 3**), we repeated the meta-analysis at that locus by incorporating genotypes of the referenced SNP(s) as additional covariate(s). For comparison of the odds ratios of the SNPs, we first selected SNPs that were shared between the meta-analyses in Japanese and Europeans. Next, we removed the SNPs located more than 1 Mb away from each of the marker SNPs in the validated rheumatoid arthritis susceptibility loci, except for in the HLA region, where we removed the SNPs located between 24,000,000 bp to 36,000,000 bp on chromosome 6 because of the existence of long-range haplotypes with rheumatoid arthritis susceptibility in this region[28]. LD pruning of the SNPs was conducted for the SNP pairs that were in LD ($r^2 \geq 0.3$) in both HapMap Phase 2 East Asian and Utah residents of Northern and Western European ancestry (CEU) populations (release 24). Correlations of the odds ratios were evaluated using R statistical software version 2.11.0.

25. Okada, Y. *et al.* Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067 (2011).
26. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
27. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
28. Okada, Y. *et al.* Contribution of a haplotype in the HLA region to anti–cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. *Arthritis Rheum.* **60**, 3582–3590 (2009).

Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes

Lucia Kato[a], Nasim A. Begum[a], A. Maxwell Burroughs[b], Tomomitsu Doi[a,1], Jun Kawai[b], Carsten O. Daub[b], Takahisa Kawaguchi[c], Fumihiko Matsuda[c], Yoshihide Hayashizaki[b], and Tasuku Honjo[a,2]

[a]Department of Immunology and Genomic Medicine and [c]The Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan; and [b]RIKEN Omics Science Center (OSC), RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045, Japan

Activation-induced cytidine deaminase (AID) is required for both somatic hypermutation and class-switch recombination in activated B cells. AID is also known to target nonimmunoglobulin genes and introduce mutations or chromosomal translocations, eventually causing tumors. To identify as-yet-unknown AID targets, we screened early AID-induced DNA breaks by using two independent genome-wide approaches. Along with known AID targets, this screen identified a set of unique genes (*SNHG3*, *MALAT1*, *BCL7A*, and *CUX1*) and confirmed that these loci accumulated mutations as frequently as *Ig* locus after AID activation. Moreover, these genes share three important characteristics with the Ig gene: translocations in tumors, repetitive sequences, and the epigenetic modification of chromatin by H3K4 trimethylation in the vicinity of cleavage sites.

deep sequencing | end labeling by biotin oligonucleotide | microarray

Activation-induced cytidine deaminase (AID) is expressed in germinal center (GC) B cells upon antigen stimulation and is essential for two types of genetic alteration in the Ig gene: class switch recombination (CSR) and somatic hypermutation (SHM), which provide the genetic basis for antibody memory (1, 2). CSR produces antibodies with different effector functions by recombination at Ig heavy chain (H) switch (S) regions, so that the μ-chain constant (Cμ) region is replaced by a downstream C_H region. SHM introduces nontemplated point mutations in the rearranged variable (V) region genes, resulting in incremented antigen receptor affinity after clonal selection (3, 4).

Functional studies on AID mutants have shown that distinct AID domains are required for SHM and CSR, although AID has a single catalytic center (cytidine deaminase motif) in the middle of the molecule. Deletions and alterations in the N-terminal region affect both the CSR and SHM activities (5). However, AID C-terminal mutants almost completely lose CSR activity but retain or even increase SHM activity (6, 7). Although C-terminally truncated AID mutants cleave both V and S regions and induce enhanced c-myc-IgH translocations, they cannot mediate CSR, suggesting that the C-terminal domain is not required for DNA cleavage but is required to correctly pair cleaved ends (8).

The DNA cleavage of targets in CSR and SHM (the S region and V region, respectively) requires their transcription (9–12). Indeed, AID-induced mutations (SHM) are generally detected in a region within 2 kb downstream of the transcription start site (TSS) (13, 14). Transcription appears to play two roles in the targeting of cleavage sites. First, transcription is associated with the epigenetic marking of the target locus, particularly by H3K4 trimethylation (H3K4me3). The histone chaperone complex FACT is required to regulate H3K4me3 in the target S region, and FACT knockdown abolishes H3K4me3 and DNA cleavage in this region (15). Second, transcription is probably required to induce non-B structures in highly repetitive sequences such as S regions (16–18), due to excessive negative supercoiling induced immediately downstream of transcription. V regions have also been shown to form stem-loop structures under these conditions

(19, 20). Non-B structure involvement has recently been reported in transcription-associated mutations in repetitive sequences such as the dinucleotide repeat hot spots or triplet repeat expansion/contractions causing Huntington's disease (17, 21, 22).

AID-dependent DNA cleavage is, in general, specific to the Ig locus. However, a number of reports have shown that AID can induce DNA cleavage in non-Ig loci. AID non-Ig targets were first demonstrated by studies on AID transgenic mice that produce numerous T lymphomas, in which vast numbers of mutations accumulate in the genes encoding the T-cell receptor, CD4, CD5, c-myc, and PIM1 (23, 24). This finding was followed by the observations that AID deficiency abolishes c-myc-Ig translocation and reduces the incidence of plasmacytoma (25, 26). AID expression is specific to activated B cells under normal conditions. However, AID expression has also been found in non-B cells, especially in cells stimulated by infection with pathogens such as human T-cell leukemia virus type 1 (HTLV1), hepatitis C virus (HCV), Epstein–Barr (EB) virus, and *Helicobacter pylori* (27–30). Based on these observations, AID is postulated to induce tumorigenesis, especially in B lymphomas and leukemias—and AID is expressed in many GC-derived human B-cell lymphomas (31–33). The prognosis of acute lymphocytic leukemia (ALL) and chronic myeloid leukemia (CML) is linked with AID expression (34, 35). It is therefore important to determine which non-Ig genes can be targeted by AID, and what features, if any, they share with Ig genes.

Several approaches have been used to explore AID non-Ig target genes in B cells. Candidate approaches involving the direct sequencing of proto-oncogenes, genes involved in translocations, or genes transcribed in normal GC B cells have shown that AID mutates several non-Ig genes, including *BCL6*, *MYC*, *PIM1*, and *PAX5* (24, 32, 36, 37). More recently, several efforts have been made to identify AID targets in a whole genome. These approaches have used chromatin immunoprecipitation (ChIP) of CSR-related proteins in combination with genome-wide tiling microarrays (ChIP-chip) or deep sequencing (ChIP-seq) on the assumption that proteins involved in CSR bind to AID targets. RPA, Nbs1, AID itself, and Spt5 have been used as marking proteins in this type of study (38–40). However, these approaches did not necessarily show that all of the protein-bound targets are cleaved or mutated by AID. There are indications that some genes identified by such approaches are not tran-

IMMUNOLOGY

scribed (39). Therefore, it is important to reexamine non-Ig AID target genes by using a different strategy.

Here, we report four AID targets, identified by a combination of unique techniques. After directly labeling the DNA breakage ends from AID-induced cleavage with a biotinylated linker, we isolated the labeled fragments with streptavidin beads and analyzed them by a combination of promoter arrays and genome-wide sequencing. The candidates identified were then confirmed by quantitative PCR (qPCR) and the actual demonstration of mutations. With these methods, we identified at least four previously unknown AID targets—*SNHG3*, *MALAT1*, *BCL7A*, and *CUX1*. We found that these targets share important characteristics with Ig genes, namely, repetitive sequences that can form non-B structures upon efficient transcription, and the accumulation of H3K4me3 histone modifications on the chromatin.

Results

AID-Induced DNA Cleavage Detected by Labeling DNA Break Ends with a Biotinylated Linker.
To detect genome-wide AID-induced DNA breaks, we used a modified in situ DNA end-labeling technique as described (8, 41) in BL2 cells, a Burkitt's lymphoma cell line that serves as an in vitro model for studying the SHM mechanism (31, 42, 43). We used the BL2 clone BL2-ΔC-AIDER, which expresses JP8Bdel, an AID mutant lacking the C-terminal 16 residues, fused with the hormone-binding domain of the estrogen receptor (ER) (JP8Bdel-ER). Tamoxifen (4-OHT) treatment induces DNA breakage in the Sμ and Sα regions but not in the Sγ region of JP8Bdel-ER-expressing CH12 cells, which switch almost exclusively from IgM to IgA (8).

BL2-ΔC-AIDER cells were treated with 4-OHT only for 3 h to minimize cell death and DNA break ends were labeled with a biotinylated linker, and the break-enriched biotinylated DNA was used as a PCR template (Fig. 1A). In agreement with previous reports (8, 42), we detected DNA breakage in the 5′ Sμ region of the IgH locus only in 4-OHT–treated cells. No breakage was detected in the *B2M* gene, which is expressed in BL2 cells but was shown not to accumulate mutations in activated B cells (Fig. 1B).

AID Targets Identified by Promoter Array and Whole Genome Sequencing.
Because SHM is normally detected close to the TSS (13, 14), biotin linker-enriched DNA fragments were analyzed by a promoter array to identify unknown AID targets. Table S1 lists the genes whose signals increased after 3 h of 4-OHT treatment, compared with untreated samples with false discovery rate (FDR) values <0.3. We also looked for genes with increased signals after 4-OHT treatment that are known to be targets of chromosomal translocation or genes that had multiple breakage peaks, and we identified >50 genes, among which we found that *BCL7A* and *CUX1* are enriched in the original breakage-enriched library by qPCR (see below). We confirmed by RT-PCR and expression array that *SNHG3*, *MALAT1*, *NIN*, *C9orf72*, *CFLAR*, *SNX25*, *BCL7A*, and *CUX1* were transcribed in BL2 cells (Table S1). Fig. S1 shows the peak signals in a 10-kb segment surrounding the breakage area of *SNHG3*, *MALAT1*, *BCL7A*, and *CUX1*. We could not map the breakage in the Ig locus because of the absence of array probes in this region.

Because the promoter array does not detect DNA fragments outside of regions containing probes, we further analyzed the breakage-enriched DNA by direct sequencing of the biotin linker-enriched library. DNA breakage sites in both control and 4-OHT–treated libraries were identified by aligning sequenced tags to the genome, and significantly enriched regions were identified by comparing the local breakage density (*SI Materials and Methods*). Regions were identified in the genes listed in Table S2. Interestingly, *SNHG3* and *MALAT1*, which were identified by the promoter array, appear at the top of the list in the genome-wide sequencing as well.
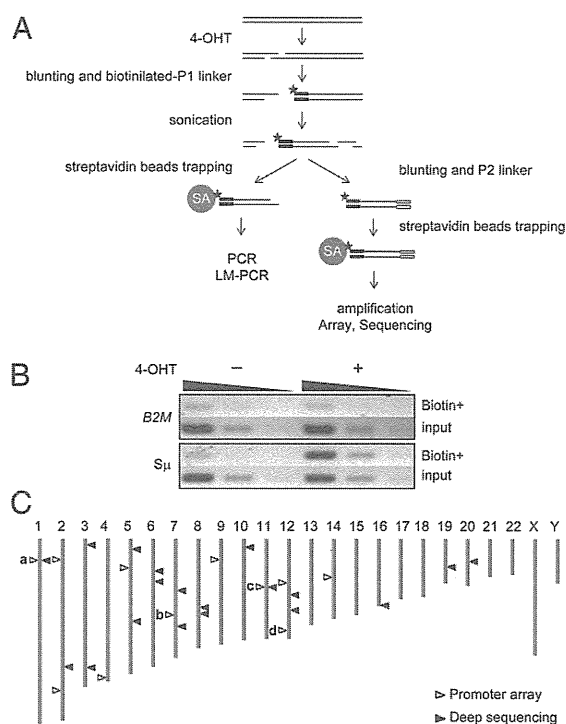

Fig. 1. (A) Schematic of the labeling technique. 4-OHT is added to activate AID, and DNA break ends are labeled in situ by biotinylated linker ligation. After genomic DNA is extracted and sonicated, biotinylated fragments are captured by streptavidin beads and used for PCR, array, or sequencing. (B) Detection of DNA breaks by PCR. BL2-ΔC-AIDER cells were treated with or without 4-OHT for 3 h, and the break ends were labeled. PCR of Sμ and *B2M* was performed with biotin-labeled DNA or input DNA by using fivefold serially diluted templates. (C) Chromosomal distribution of AID targets. a, *SNHG3*; b, *CUX1*; c, *MALAT1*; d, *BCL7A*. White arrowhead, promoter array (FDR < 0.3 plus *BCL7A* and *CUX1*); black arrowhead, whole genome sequencing (FDR < 0.01 and/or remarkable numbers of *P* value clusters).

Fig. 1C shows the chromosomal distribution of AID target candidates identified by promoter array or whole-genome sequencing. Breakage seemed to be distributed through the genome without any apparent bias. Surprisingly, of the 29 candidates identified by whole-genome sequencing with strict statistical parameters, only two matched candidates obtained from the promoter array. This discrepancy might be explained in part because most of the breakage-rich regions detected by whole genome sequencing are located in regions that do not contain promoter array probes.

Results may also be limited because of possible bias by PCR amplification of the primary library for microarray and whole-genome sequencing, which could affect the relative genome coverage. To avoid this bias, we relied on the original library and confirmed all candidates by qPCR.

qPCR Analyses of Linker Libraries.
To confirm the AID-induced breakage candidates detected by the promoter array and whole-genome sequencing, we used qPCR assays with gene-specific primers to amplify the vicinity of the identified breakage regions in biotin linker-enriched DNA from cells treated with 4-OHT for 3 h (Fig. 2). We examined whether candidate genes were enriched in the 4-OHT–treated DNA library compared with the nontreated library. Among the 29 candidates identified by whole-genome sequencing, only *SNHG3* and *MALAT1* were strongly enriched (*P* < 0.0001 and *P* < 0.001, respectively). Besides these, *BCL7A*, *CUX1*, and *CFLAR*, which were picked up only by the
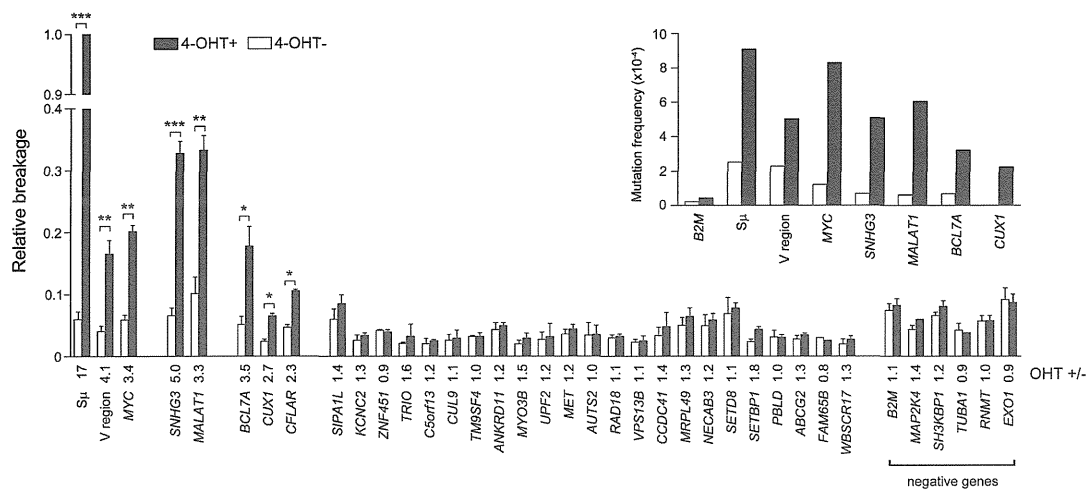
199

Fig. 2. qPCR measurement of DNA breaks. Break signals are presented relative to Sμ. SD values were derived from at least three independent experiments, and P values were calculated by a two-tailed t test. $*P < 0.01$, $**P < 0.001$, $***P < 0.0001$. Numbers below the x axis indicate the ratio between samples treated and not treated with 4-OHT. (Inset) Mutation analysis of genes with significantly increased break signals after AID activation. Cells were treated with or without 4-OHT for 24 h. Only unique mutations were counted. Detailed mutation profiles can be found in Fig. S2 and Table S3.

promoter array, also showed significant enrichment ($P < 0.01$) in the 4-OHT–treated library.

We also confirmed that the Sμ and V regions in BL2 cells were cleaved, because they were enriched in the 4-OHT–treated library. Although *MYC*, which is translocated in an AID-dependent manner in human Burkitt's lymphoma (44), was not identified by either promoter array or whole-genome sequencing, qPCR of the 4-OHT–treated samples clearly revealed *MYC* gene enrichment (Fig. 2). The difference in cleavage detection between the direct candidate qPCR and genome-wide arrays and sequencing suggests that the amplification step required for microarray and whole-genome sequencing methods may introduce bias, either for or against many genes. In the case of sequencing, this bias can lead to low mapping coverage of certain regions, hampering efforts to identify significant enrichment. Therefore, we cannot exclude genes that were not identified by the present methods from being AID targets.

AID Targets Accumulate Somatic Mutations near Cleavage Sites. To test whether the newly identified target genes are mutated upon AID activation, we treated BL2-ΔC-AIDER cells with 4-OHT for 24 h and sequenced regions of ≈600 bp around each area with abundant breakage (Fig. S2 and Table S3). Mutations increased in all of the qPCR-confirmed AID target genes after 4-OHT treatment (Fig. 2, *Inset*), with mutation frequencies ranging from 6.1×10^{-4} for *MALAT1* to 2.2×10^{-4} for *CUX*. These frequencies are comparable to those of the V region (5.0×10^{-4}), the Sμ region (9.1×10^{-4}), and the *MYC* gene (8.3×10^{-4}), and are far higher than that of the control *B2M* gene (4.3×10^{-5}). We also detected mutations in the *CFLAR* gene; however, the mutation frequency (9.2×10^{-5}) was not as high as other AID target genes, although mutations increased significantly in 4-OHT–treated sample ($P = 0.004$) (Table S3).

To compare the distribution profiles of mutated bases and AID-induced DNA breaks in the biotin linker-enriched DNA, we mapped the linker positions by performing ligation-mediated (LM)-PCR with the linker primer and gene-specific primers. These PCR fragments were subsequently cloned and sequenced. Break ends identified by the linker were plotted, together with mutation positions (Fig. 3 and Fig. S2). The results clearly showed that the DNA cleavage marks (biotin linker) were closely associated with mutations, indicating that the DNA cleavage

sites identified are functionally relevant to SHM by AID. We used RT-PCR and expression arrays to confirm that the regions
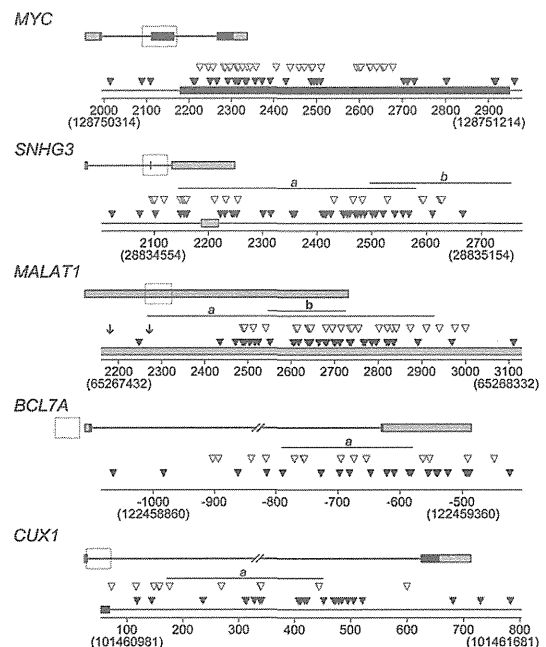

Fig. 3. Somatic mutations and breakpoint distribution in AID target loci. Mutations (open triangles) and breakpoints (filled triangles) detected by LM-PCR (Fig. S2) were plotted on the respective genomic sequences. The top scheme represents exons (rectangles) and introns (bars). Genomic loci are shown in untranslated and translated sequences (gray and black boxes, respectively). The horizontal lines a and b represent breakage regions identified by promoter array and sequencing, respectively. Regions outlined by dotted boxes are shown in more detail below each genomic locus. For the *MALAT1* locus, the translocation breakpoints reported by Davis et al. (45) are represented by arrows. x axis numbers indicate base positions according to RefSeq: NM_002467 (*MYC*), NR_002909 (*SNHG3*), NR_002819 (*MALAT1*), NM_020993 (*BCL7A*), and NM_181552 (*CUX1*). Numbers in parentheses indicate the corresponding base position according to hg19 assembly.

200

where DNA cleavage and mutations were identified are transcribed (Tables S1 and S2).

Repetitive Sequences Surround the Breakage Regions of Unique Targets.

We next examined common features among the AID targets. Although SHM has been reported to prefer the RGYW-WRCY motif (46), we could not find any enrichment of this motif among the break sites in the newly identified targets. It was recently reported that mutations are introduced in regions with sequences prone to forming non-B DNA structure, including tandem repeats, palindromes, and inverted repeats (17, 18). The S region, MYC, and V region genes contain sequences prone to forming non-B structure (19, 20, 47, 48). We used REPFIND, a program that identifies clustered, nonrandom short repeats in a given nucleotide sequence, to search the vicinity of identified breakage regions for sequences prone to forming non-B structure. For each repeat cluster, a P value is calculated indicating the probability of finding such a repeat cluster randomly (a P value of 1×10^{-5} means that such a concentration of that particular repeat occurs an average of once in 100,000 bp by chance) (49). Curiously, we found that various types of repeat sequences cluster in the vicinity of cleaved sites in the newly identified AID target genes. In the MALAT1 locus, the region within 2 kb surrounding the breakage peaks was rich in clustered short repeat motifs such as GAAG, GCC, GAA, CCG, AAG, GAAGA, and TTAA (Fig. 4). Repeat clusters were also found near the cleavage sites of the SNHG3, BCL7A, and CUX1 loci. (Fig. S3). In all cases, the probability of the appearance of these repeats was far below random ($P < 1 \times 10^{-8}$).

H3K4me3 at Cleavage Sites.

It was recently shown that S region transcription alone is not sufficient for CSR; specific histone posttranslational modification marks, especially H3K4me3, are required. H3K4me3 depletion strongly inhibits CSR and DNA cleavage in the Sμ and Sα regions (15). We thus asked whether the V region and the newly identified AID targets also carry H3K4me3 marks around the cleavage regions. ChIP analysis showed that both the V region and MALAT1 locus were abundantly marked by H3K4me3 (Fig. 5). Furthermore, the H3K4me3 distribution profiles corresponded well to the somatic mutation distribution in the rearranged V region and to the breakage signal
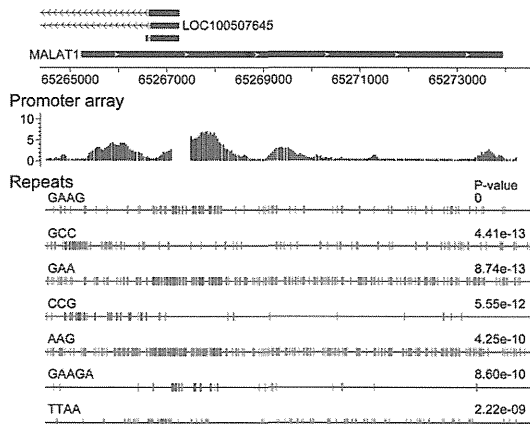
Fig. 4. Repeat sequences surrounding the breakage region in the MALAT1 gene. (Top) Representation of a 10-kb segment surrounding the MALAT1 locus. x axis numbers represent base positions according to hg19 assembly. (Middle) Breakage signal distribution detected by promoter array. Regions without bars do not have array probes. (Bottom) REPFIND analysis showing significant repeat clusters in the MALAT1 locus. Motifs depicted as small, colored, vertical bars indicate the cluster with the most significant P value; individual repeats are separated by different colors.
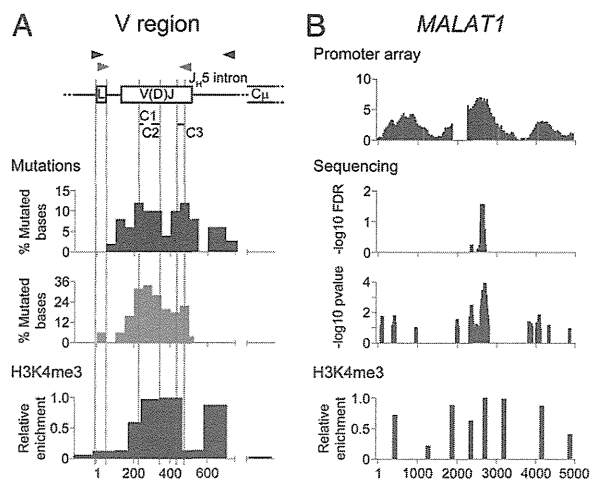
Fig. 5. H3K4me3 distribution in the IgH V region and in the MALAT1 gene. (A Top: Representation of the rearranged IgH V region of BL2 cells. Black and gray arrowheads represent the position of primers used for the mutation analysis shown in Bottom (graphs in black and gray, respectively). L, leader; C1, CDR1; C2, CDR2; C3, CDR3. (A Middle) Somatic mutation distribution, represented as the percentage of mutated bases per 50 bp sequenced. Graph in black: mutations from Fig. 2, Inset. Graph in gray: mutations reported by Denepoux et al. (50). (Bottom) ChIP assay using an anti-H3K4me3 antibody. x axis numbers indicate the nucleotide position relative to the first V-gene ATG. (B) MALAT1 locus. From top to bottom: Breakage signal distribution detected by promoter array (regions without bars do not have array probes); FDR regions by sequencing; P value peaks by sequencing; ChIP assay using an H3K4me3 antibody. x axis numbers indicate base positions according to RefSeq NR_002819.

distribution observed by both the promoter array and whole genome sequencing in MALAT1 (Fig. 5 A and B). Mutations identified in MALAT1 overlapped with DNA cleavage signals and H3K4me3 marks (Figs. 3 and 5B). We examined the H3K4me3 pattern of other AID targets by using publicly available ENCODE ChIP-seq data for the B-lymphoblastoid cell line GM12878 (51). As expected, all of them, except for BCL7A, were highly abundant in H3K4me3 marks overlapping nicely with cleavage sites (Fig. S4). H3K4me3 might be absent at the BCL7A locus in GM12878 cells because it is an inducible gene expressed in BL2 cells, but not in the GM12878 cell line (52). We thus conclude that the newly identified AID targets share both cis and trans marks for AID targeting—non-B structure and H3K4me3, respectively (15, 16).

Discussion

Identified AID Targets Accumulate High-Frequency Mutations. We explored AID targets by combining three different strategies: promoter array, whole genome sequencing, and candidate qPCR in a library containing biotinylated linker-labeled cleaved ends. With these strong criteria, we were able to identify four unique AID targets: SNHG3, MALAT1, BCL7A, and CUX1. All of these candidates were further confirmed to accumulate mutations. These candidates are thus strong AID cleavage targets; however, these genes represent only very efficient AID targets. The use of the biotinylated linker, which efficiently identifies double-strand breakage with close, staggered nicks on opposite strands, may not detect scattered nicks efficiently, and this may limit identification to targets that are efficiently and specifically cleaved within 3 h of AID activation.

Some well-described SHM target genes, including MYC, BCL6, PAX5, RHOH, and PIM1, were not detected by either the promoter array or whole genome sequencing. We used qPCR to test whether these genes were enriched in the biotin-labeled

DNA library, but only *MYC* was enriched in the 4-OHT–treated sample (Fig. 2). These genes have been found to be mutated in memory and GC B cells as well as lymphoma cells (24, 32, 36, 37), cells that are expected to be chronically exposed to AID. In addition, the mutation accumulation in tumor cells depends on selection. In contrast, in our study, we exposed BL2 cells to a short treatment (3 h) of 4-OHT, to increase the chance of detecting only efficiently targeted loci. In fact, none of the genes above mentioned mutated more than 1/20th of the 3′ J_H locus even in 6-mo-old Peyer's patch B cells (36).

The unique AID targets accumulate mutations at comparable frequencies with the *Ig* and *MYC* genes. We found that the mutation and cleavage sites are located in similar areas. The results indicate that the cleavage and mutation sites are linked, but not necessarily identical. This observation is consistent with the prediction that SHM is incorporated during the repair phase by error-prone polymerases (53). We confirmed that all of the newly identified AID targets were highly transcribed in BL2 cells. Although the breakage signal detected at the *BCL7A* locus was ≈800 bp upstream of the TSS, we detected both sense and antisense transcripts in this region.

Unique AID Targets also Translocate. Furthermore, it is important to stress that all of these unique candidates have been shown to be the targets of chromosomal translocation in neoplastic cells as shown for the *Ig* locus and *MYC* gene. MALAT1 is overexpressed in several cancers and was reported to be involved in regulating alternative splicing (54). The *MALAT1* locus has been found to harbor chromosomal translocation breakpoints associated with cancer (45, 55) and, interestingly, two reported translocation breakpoints are close to or within the breakage region identified in the present study (Fig. 3). *SNHG3*, a host gene for small nucleolar RNAs (56), is also reported to be involved in translocation, and although the exact position of the translocation breakpoint has not been reported, we can speculate that it is located in the second intron of *SNHG3* because the detected fusion transcript joins the second exon of *SNHG3* with the exon of the 3′ partner gene (57). *BCL7A* and *CUX1* have also been reported to bear chromosomal translocations; however, these translocation breakpoints occur far from the breakage regions identified in this study (58, 59).

Abundant Repetitive Sequences in AID Targets. To identify common features of AID targets, we compared the *MYC*, *SNHG3*, *MALAT1*, *CUX1*, and *BCL7A* genes with the Ig gene locus (the V_H gene and the Sμ region). Sequence analysis identified abundant repetitive sequences surrounding the cleaved regions of AID targets. A typical example is *MALAT1* (Fig. 4): The GAAG, GCC, GAA, CCG, AAG, GAAGA, and TTAA repeats are highly abundant within 2 kb surrounding the break peaks, which also overlap with actual mutation sites. In the *SNHG3* locus, less frequent but longer repeats—GGATTACAG, TTT-TTGTATTT, ATTACAGGC, GCCTC, and TTTTTGTA—are clustered in the proximity of cleavage sites (Fig. S3*A*). *BCL7A* and *CUX1* have GC-rich repeats, such as CGCG, CCGCG, CCCG, and CGGCG (Fig. S2 *B* and *C*). The *MYC* gene, the V region, and the S region are already known to have repetitive sequences or inverted repeats that can form non-B structure when the target is actively transcribed and under an excessive negative superhelical condition (19, 20, 47, 48).

H3K4me3 Marks in AID Targets. Chromatin modifications are also involved in AID targeting. We showed that H3K4 methylation, specifically trimethylation, is critical for DNA cleavage in the S region (15), although Odegard et al. (60) showed that the H3K4 dimethylation (H3K4me2) pattern is similar among VJλ1, Cλ1, and Eλ3-1 and concluded that H3K4me2 is not correlated with SHM. Association of H3K4me3 with the MYC locus was also

reported (38). Therefore, we tested whether H3K4me3 modification is also associated with the V region and the unique loci. SHM in V regions typically targets the whole coding V-region segment and extends to its 5′ and 3′ flanking regions. Mutation frequencies rise sharply ≈100 bp downstream of the TSS (at the middle of the leader intron), peak in V(D)J, and then gradually decrease after the immediate 3′ flanking region, becoming undetectable over a distance of ≈1 kb from the rearranged J (61). It is striking that the H3K4me3 profile follows the exact same tendency as SHM distribution in the V region (Fig. 5*A*). H3K4me3 is scarce in the leader exon and intron but present in the highly mutated portion of the V(D)J exon. We also observed that H3K4me3 distribution at the *MALAT1* locus corresponded well with the breakage signal distribution detected by both the promoter array and whole genome sequencing (Fig. 5*B* and Fig. S3*A*). The H3K4me3 pattern of other AID targets also overlaps with cleavage sites (Fig. S3 *B–D*). Strikingly, we observed a strong H3K4me3 peak in the 5′ region of the *CUX1* gene (Fig. S4*D*), which does not contain microarray probes. We confirmed that this region also accumulates mutations after 4-OHT treatment (Table S3). It would be interesting to check whether H3K4me3 depletion can decrease AID-induced breaks and mutations in the newly identified AID targets.

We thus conclude that all of these genes, *SNHG3*, *MALAT1*, *BCL7A*, and *CUX1*, share unique characteristics that are required for AID targeting: non-B structure as the *cis* element and the H3K4me3 histone modification as the *trans* mark.

Materials and Methods

Labeling of DNA Break Ends by a Biotinylated Linker. The biotin-labeled DNA break assay was performed as described (8) with slight modifications. After nuclear permeabilization, BL2 cells were washed with cold PBS and resuspended in 1× T4 DNA polymerase buffer. Blunting was performed by using T4 DNA Polymerase (Takara). After washing with cold PBS, 4 μL of T4 DNA Ligase (Takara) and 13.4 μL of an annealed biotinylated P1 linker were added, and the cells were incubated overnight at 16 °C. Genomic DNA was purified by phenol:chloroform extraction.

PCR, Real-Time PCR, and LM-PCR. Biotinylated genomic DNA (10 μg) was sonicated (Covaris) and incubated with 10 μL of M-270 Dynabeads (Invitrogen) for 15 min at room temperature. After washing, the beads were resuspended in 15 μL of TE buffer and used as a PCR template. PCR was initiated by denaturing for 5 min at 95 °C followed by 25 cycles (95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s) and a final extension at 72 °C for 5 min. SYBR Green Master Mix (Applied Biosystems) was used for real-time PCR.

For LM-PCR, we used a template of 1 μL of beads in a two-round PCR by using linker primer (P1-LM) and gene-specific primers. First-round PCR was initiated by nick translation (72 °C for 20 min), followed by denaturing (95 °C for 5 min), 25 cycles (95 °C for 15 s, 65 °C for 15 s, and 70 °C for 1 min), and a final extension (70 °C for 5 min). Second-round PCR included denaturing (95 °C for 5 min), 20 cycles (95 °C for 15 s, 65 °C for 15 s, and 70 °C for 1 min), and a final extension (70 °C for 7 min). The PCR fragments were purified, cloned with the pGEM-T Easy Vector System (Promega), and sequenced with the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). Primers sequences are provided in Table S4–S7.

DNA Preparation for Microarray and SOLiD Sequencing. After sonication of biotin-labeled genomic DNA, sheared ends were blunted by adding T4 DNA polymerase for 30 min at room temperature. DNA was purified by using the PureLink PCR purification Kit (Invitrogen), P2-annealed linker was ligated overnight at 16 °C, DNA was incubated with Dynabeads as described above, and the beads were used for global amplification by following the SOLiD protocol (Applied Biosystems). A summary of general features of the sequenced libraries can be found in Fig. S5 and Table S8.

Accession Codes. Gene Expression Omnibus: microarray data, GSE32027; DNA Data Bank of Japan: sequencing data, DRA000450.

Other material and methods are provided in *SI Materials and Methods*.

1. Muramatsu M, et al. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102:553–563.
2. Revy P, et al. (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102:565–575.
3. Honjo T, Kinoshita K, Muramatsu M (2002) Molecular mechanism of class switch recombination: Linkage with somatic hypermutation. *Annu Rev Immunol* 20:165–196.
4. Teng G, Papavasiliou FN (2007) Immunoglobulin somatic hypermutation. *Annu Rev Genet* 41:107–120.
5. Shinkura R, et al. (2004) Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* 5:707–712.
6. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC (2003) C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 12:501–508.
7. Ta VT, et al. (2003) AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* 4:843–848.
8. Doi T, et al. (2009) The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proc Natl Acad Sci USA* 106:2758–2763.
9. Jung S, Rajewsky K, Radbruch A (1993) Shutdown of class switch recombination by deletion of a switch region control element. *Science* 259:984–987.
10. Peters A, Storb U (1996) Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* 4:57–65.
11. Betz AG, et al. (1994) Elements regulating somatic hypermutation of an immunoglobulin kappa gene: Critical role for the intron enhancer/matrix attachment region. *Cell* 77:239–248.
12. Zhang J, Bottaro A, Li S, Stewart V, Alt FW (1993) A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. *EMBO J* 12:3529–3537.
13. Hackett J, Jr., Rogerson BJ, O'Brien RL, Storb U (1990) Analysis of somatic mutations in kappa transgenes. *J Exp Med* 172:131–137.
14. O'Brien RL, Brinster RL, Storb U (1987) Somatic hypermutation of an immunoglobulin transgene in kappa transgenic mice. *Nature* 326:405–409.
15. Stanlie A, Aida M, Muramatsu M, Honjo T, Begum NA (2010) Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proc Natl Acad Sci USA* 107:22190–22195.
16. Kobayashi M, et al. (2009) AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc Natl Acad Sci USA* 106:22375–22380.
17. Hubert L, Jr., Lin Y, Dion V, Wilson JH (2011) Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. *Mol Cell Biol* 31:3105–3112.
18. Zhao J, Bacolla A, Wang G, Vasquez KM (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci* 67:43–62.
19. Rogozin IB, Solovyov VV, Kolchanov NA (1991) Somatic hypermutagenesis in immunoglobulin genes. I. Correlation between somatic mutations and repeats. Somatic mutation properties and clonal selection. *Biochim Biophys Acta* 1089:175–182.
20. Wright BE, Schmidt KH, Minnick MF, Davis N (2008) I. VH gene transcription creates stabilized secondary structures for coordinated mutagenesis during somatic hypermutation. *Mol Immunol* 45:3589–3599.
21. Lippert MJ, et al. (2011) Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc Natl Acad Sci USA* 108:698–703.
22. Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S (2011) Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 108:692–697.
23. Okazaki IM, et al. (2003) Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 197:1173–1181.
24. Kotani A, et al. (2005) A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc Natl Acad Sci USA* 102:4506–4511.
25. Ramiro AR, et al. (2004) AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 118:431–438.
26. Takizawa M, et al. (2008) AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med* 205:1949–1957.
27. Ishikawa C, Nakachi S, Senba M, Sugai M, Mori N (2011) Activation of AID by human T-cell leukemia virus Tax oncoprotein and the possible role of its constitutive expression in ATL genesis. *Carcinogenesis* 32:110–119.
28. Machida K, et al. (2004) Hepatitis C virus induces a mutator phenotype: Enhanced mutations of immunoglobulin and protooncogenes. *Proc Natl Acad Sci USA* 101:4262–4267.
29. Epeldegui M, Hung YP, McQuay A, Ambinder RF, Martínez-Maza O (2007) Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. *Mol Immunol* 44:934–942.
30. Matsumoto Y, et al. (2007) Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* 13:470–476.
31. Faili A, et al. (2002) AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* 3:815–821.
32. Pasqualucci L, et al. (2001) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412:341–346.
33. Pasqualucci L, et al. (2004) Expression of the AID protein in normal and neoplastic B cells. *Blood* 104:3318–3325.
34. Feldhahn N, et al. (2007) Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. *J Exp Med* 204:1157–1166.
35. Leuenberger M, et al. (2010) AID protein expression in chronic lymphocytic leukemia/small lymphocytic lymphoma is associated with poor prognosis and complex genetic alterations. *Mod Pathol* 23:177–186.
36. Liu M, et al. (2008) Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451:841–845.
37. Shen HM, Peters A, Baron B, Zhu X, Storb U (1998) Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* 280:1750–1752.
38. Yamane A, et al. (2011) Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 12:62–69.
39. Staszewski O, et al. (2011) Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol Cell* 41:232–242.
40. Pavri R, et al. (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143:122–133.
41. Ju BG, et al. (2006) A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* 312:1798–1802.
42. Nagaoka H, Ito S, Muramatsu M, Nakata M, Honjo T (2005) DNA cleavage in immunoglobulin somatic hypermutation depends on de novo protein synthesis but not on uracil DNA glycosylase. *Proc Natl Acad Sci USA* 102:2022–2027.
43. Woo CJ, Martin A, Scharff MD (2003) Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. *Immunity* 19:479–489.
44. Dalla-Favera R, et al. (1982) Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc Natl Acad Sci USA* 79:7824–7827.
45. Davis IJ, et al. (2003) Cloning of an Alpha-TFEB fusion in renal tumors harboring the t (6;11)(p21;q13) chromosome translocation. *Proc Natl Acad Sci USA* 100:6051–6056.
46. Rogozin IB, Kolchanov NA (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta* 1171:11–18.
47. Tashiro J, Kinoshita K, Honjo T (2001) Palindromic but not G-rich sequences are targets of class switch recombination. *Int Immunol* 13:495–505.
48. Michelotti GA, et al. (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol Cell Biol* 16:2656–2669.
49. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO (2002) A ubiquitous and conserved signal for RNA localization in chordates. *Curr Biol* 12:1756–1761.
50. Denépoux S, et al. (1997) Induction of somatic mutation in a human B cell line in vitro. *Immunity* 6:35–46.
51. Birney E, et al.; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
52. Ernst J, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49.
53. Faili A, et al. (2004) DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. *J Exp Med* 199:265–270.
54. Tripathi V, et al. (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39:925–938.
55. Rajaram V, Knezevich S, Bove KE, Perry A, Pfeifer JD (2007) DNA sequence of the translocation breakpoints in undifferentiated embryonal sarcoma arising in mesenchymal hamartoma of the liver harboring the t(11;19)(q11;q13.4) translocation. *Genes Chromosomes Cancer* 46:508–513.
56. Pelczar P, Filipowicz W (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol Cell Biol* 18:4509–4518.
57. Levin JZ, et al. (2009) Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol* 10:R115.
58. Zani VJ, et al. (1996) Molecular cloning of complex chromosomal translocation t(8;14; 12)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (BCL7A) with homology to caldesmon. *Blood* 87:3124–3134.
59. Wasag B, Lierman E, Meeus P, Cools J, Vandenberghe P (2011) The kinase inhibitor TKI258 is active against the novel CUX1-FGFR1 fusion detected in a patient with T-lymphoblastic leukemia/lymphoma and t(7;8)(q22;p11). *Haematologica* 96:922–926.
60. Odegard VH, Kim ST, Anderson SM, Shlomchik MJ, Schatz DG (2005) Histone modifications associated with somatic hypermutation. *Immunity* 23:101–110.
61. Lebecque SG, Gearhart PJ (1990) Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J Exp Med* 172:1717–1727.

A Genome-Wide Association Study Identified *AFF1* as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese

Yukinori Okada[1,2,3⦾], Kenichi Shimane[1,2⦾], Yuta Kochi[1,2⦾*], Tomoko Tahira[4], Akari Suzuki[1], Koichiro Higasa[3], Atsushi Takahashi[3], Tetsuya Horita[5], Tatsuya Atsumi[5], Tomonori Ishii[6], Akiko Okamoto[2], Keishi Fujio[2], Michito Hirakata[7], Hirofumi Amano[8], Yuya Kondo[9], Satoshi Ito[9], Kazuki Takada[10], Akio Mimori[11], Kazuyoshi Saito[12], Makoto Kamachi[13], Yasushi Kawaguchi[14], Katsunori Ikari[14], Osman Wael Mohammed[15], Koichi Matsuda[15], Chikashi Terao[16,17], Koichiro Ohmura[16], Keiko Myouzen[1], Naoya Hosono[18], Tatsuhiko Tsunoda[19], Norihiro Nishimoto[20], Tsuneyo Mimori[16], Fumihiko Matsuda[17], Yoshiya Tanaka[12], Takayuki Sumida[9], Hisashi Yamanaka[14], Yoshinari Takasaki[8], Takao Koike[5], Takahiko Horiuchi[21], Kenshi Hayashi[4], Michiaki Kubo[18], Naoyuki Kamatani[3], Ryo Yamada[1,17], Yusuke Nakamura[15], Kazuhiko Yamamoto[1,2]

1 Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan, 2 Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, 3 Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan, 4 Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan, 5 Department of Medicine II, Hokkaido University Graduate School of Medicine, Sapporo, Japan, 6 Department of Hematology and Rheumatology, Tohoku University Graduate School of Medicine, Sendai, Japan, 7 Division of Rheumatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan, 8 Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine, Tokyo, Japan, 9 Division of Clinical Immunology, Doctoral Program in Clinical Sciences, Graduate School of Comprehensive Human Science, University of Tsukuba, Tsukuba, Japan, 10 Departments of Medicine and Rheumatology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan, 11 Division of Rheumatic Diseases, National Center for Global Health and Medicine, Tokyo, Japan, 12 First Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan, 13 Department of Immunology and Rheumatology, Unit of Translational Medicine, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan, 14 Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, 15 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, 16 Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, 17 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, 18 Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan, 19 Laboratory for Medical Informatics, CGM, RIKEN, Yokohama, Japan, 20 Laboratory of Immune Regulation, Wakayama Medical University, Wakayama, Japan, 21 Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan

Abstract

Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Although recent genome-wide association studies (GWAS) have contributed to discovery of SLE susceptibility genes, few studies has been performed in Asian populations. Here, we report a GWAS for SLE examining 891 SLE cases and 3,384 controls and multi-stage replication studies examining 1,387 SLE cases and 28,564 controls in Japanese subjects. Considering that expression quantitative trait loci (eQTLs) have been implicated in genetic risks for autoimmune diseases, we integrated an eQTL study into the results of the GWAS. We observed enrichments of cis-eQTL positive loci among the known SLE susceptibility loci (30.8%) compared to the genome-wide SNPs (6.9%). In addition, we identified a novel association of a variant in the AF4/FMR2 family, member 1 (*AFF1*) gene at 4q21 with SLE susceptibility (rs340630; $P = 8.3 \times 10^{-9}$, odds ratio = 1.21). The risk A allele of rs340630 demonstrated a cis-eQTL effect on the *AFF1* transcript with enhanced expression levels ($P < 0.05$). As *AFF1* transcripts were prominently expressed in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes, up-regulation of *AFF1* may cause the abnormality in these lymphocytes, leading to disease onset.

Author Summary

Although recent genome-wide association study (GWAS) approaches have successfully contributed to disease gene discovery, many susceptibility loci are known to be still uncaptured due to strict significance threshold for multiple hypothesis testing. Therefore, prioritization of GWAS results by incorporating additional information is recommended. Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Considering that abnormalities in B cell activity play essential roles in SLE, prioritization based on an expression quantitative trait loci (eQTLs) study for B cells would be a promising approach. In this study, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS and identified *AFF1* as a novel SLE susceptibility loci. We also confirmed cis-regulatory effect of the locus on the *AFF1* transcript. Our study would be one of the initial successes for detecting novel genetic locus using the eQTL study, and it should contribute to our understanding of the genetic loci being uncaptured by standard GWAS approaches.

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by autoantibody production, complement activation, and multi-organ damage [1]. Familial aggregation demonstrates that both genetic and environmental factors play a role in pathogenesis of SLE [2]. Genetic studies using candidate gene-approaches, and recently, genome-wide association studies (GWAS), have uncovered more than 25 SLE susceptibility genes, including *HLA-DRB1*, *IRF5*, *STAT4*, *ITGAM*, *BLK*, *TNFAIP3*, and others [3–18]. However, most of these studies were conducted in European populations [3–13,15,17], and few studies have been conducted in Asian populations [14,16,18]. Since the epidemiology of SLE has demonstrated that the prevalence of disease substantially differs among populations, genetic backgrounds of SLE should be also heterogeneous across populations [19,20]. Therefore, additional studies in Asians might provide novel insights. It is of note that GWAS for SLE in Chinese populations identified novel loci that had not been detected in Europeans, such as *ETS1*, *IKZF1*, and *WDFY4* [14,16].

Another issue raised by the previous GWASs for complex diseases is that many susceptibility loci still remained uncaptured, owing to its strict significance threshold for multiple hypothesis testing [21]. In SLE, for example, the 26 risk loci identified by the previous GWAS explained only an estimated 8% of the total genetic susceptibility to the disease [15]. Therefore, it is still important to examine the sub-loci of GWAS, in order to reveal the entire picture of genetic etiology. To effectively explore these uncaptured loci, prioritization of GWAS results by incorporating additional information implicated in the disease pathophysiology is recommended [22,23]. Considering that abnormalities in B cell activity play essential roles in SLE [1] and that expression quantitative trait loci (eQTL) have been implicated to comprise approximately a half of genetic risks for autoimmune diseases [24], prioritization based on an eQTL study for B cells would be a promising approach for SLE [25]. Moreover, an eQTL itself assures the presence of functional variant(s) that regulate gene expression. Thus, eQTL increases the prior probability of the presence of disease-causal variant(s) in the locus more effectively

and unbiasedly, compared to other knowledge-based prioritizations such as gene pathway analysis [24].

Here, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS, which effectively enabled to detect a novel SLE susceptibility locus.

Results

GWAS for SLE

In the GWAS, 891 SLE cases and 3,384 controls in Japanese subjects were genotyped over 550,000 single nucleotide polymorphism (SNP) markers (Table S1, S2 and Figure 1). We applied stringent quality control (QC) criteria and evaluated associations of 430,797 autosomal SNPs, as previously described [26]. No substantial population stratification was demonstrated through principal component analysis (Figure S1) or a Quantile–Quantile plot of *P*-values (inflation factor, $\lambda_{GC} = 1.088$, Figure S2), suggesting homogenous ancestries of our study population [27].

We identified significant associations in six chromosomal loci that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (Table 1 and Figure 2A). These loci have been reported to be associated with SLE susceptibility (*STAT4*, *TNFAIP3*, *HIP1*, *BLK*, *ETS1*, and the HLA region) [3–18]. We also observed significant replications in 17 of the previously reported SLE susceptibility loci [3–18] ($\alpha = 0.01$; Table 2). Of these, significant replications were enriched in the loci identified through the studies in Asian populations (80%; 8 of the 10 loci), including *RASGRP3*, *IKZF1*, *HIP1*, *WDFY4*, intergenic region at 11q23, *ETS1*, *SLC15A4*, *ELF1*, and *HIC2-UBE2L3* [14,16,18], compared to those in European populations (56.3%; 9 of the 16 loci) [3–13,15,17].

Incorporation of eQTL study into GWAS results

For the selection of SNPs incorporated in the replication studies of the potential association signals, we evaluated cis-eQTL effects of the SNPs using publically available gene expression data [28], and prioritized the results of the GWAS. After applying QC criteria, we evaluated the expression levels of 19,047 probes assayed in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals [29] using Illumina's human whole-genome expression array (WG-6 version 1) [28]. For each of the SNPs included in our GWAS, probes located within ±300 kbp regions were focused on as cis-eQTLs (average 4.93 probes per SNP). We denoted the SNPs which exhibited significant associations with expression levels of any of the corresponding cis-eQTLs as eQTL positive (false discovery rate (FDR) Q-values<0.2). We observed enrichments of eQTL positive loci among the SLE susceptibility loci (30.8%; 8 of the 26 evaluated loci) including a well-known eQTL gene of *BLK* [11,25] (Table 2), compared to the genome-wide SNPs (6.9%) and compared even to the SNPs in the vicinity of expressed loci (among the SNPs located within ±10 kbp of probes used for the expression analysis, 13.1% were eQTL positive; Table S3).

By prioritizing the results of the GWAS using the eQTL study, we selected 57 SNPs from 1,207 SNPs that satisfied $P < 1.0 \times 10^{-3}$ in the GWAS. We subsequently referred the associations of the selected SNPs using the results of the concurrent genome-wide scan for SLE in an independent Japanese population (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009). In the scan, 447 SLE cases and 680 controls of Japanese origin were evaluated using a pooled DNA approach [30]. We selected SNPs if any association signals were observed in the neighboring SNPs of the
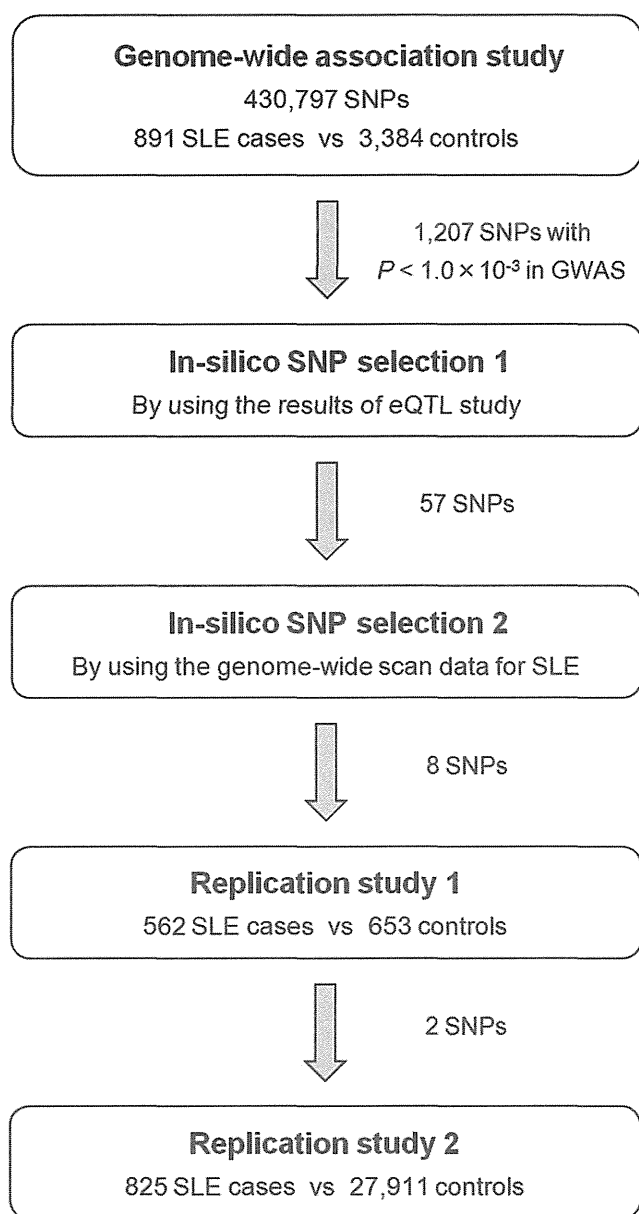
205

```
┌─────────────────────────────────────────────┐
│        Genome-wide association study          │
│              430,797 SNPs                     │
│      891 SLE cases  vs  3,384 controls        │
└─────────────────────────────────────────────┘
                    ⇓  1,207 SNPs with
                       P < 1.0 × 10⁻³ in GWAS
┌─────────────────────────────────────────────┐
│          In-silico SNP selection 1            │
│      By using the results of eQTL study       │
└─────────────────────────────────────────────┘
                    ⇓  57 SNPs
┌─────────────────────────────────────────────┐
│          In-silico SNP selection 2            │
│   By using the genome-wide scan data for SLE  │
└─────────────────────────────────────────────┘
                    ⇓  8 SNPs
┌─────────────────────────────────────────────┐
│             Replication study 1               │
│        562 SLE cases  vs  653 controls        │
└─────────────────────────────────────────────┘
                    ⇓  2 SNPs
┌─────────────────────────────────────────────┐
│             Replication study 2               │
│      825 SLE cases  vs  27,911 controls       │
└─────────────────────────────────────────────┘
```

Figure 1. Design of the GWAS and multi-stage replication studies for SLE in Japanese subjects. A total of 2,278 SLE cases and 31,948 controls were enrolled. The clinical characteristics of the subjects are summarized in Table S1 and S2. Details of the genome-wide scan data for SLE referenced in the *in silico* SNP selection 2 are described elsewhere (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009).
doi:10.1371/journal.pgen.1002455.g001

pooled analysis. As a result, 8 SNPs remained for further investigation (Table S4).

Replication studies and identification of *AFF1*

Then, we performed two-stage replication studies using independent SLE cohorts for Japanese subjects (cohort 1 with 562 SLE cases and 653 controls, and cohort 2 with 825 SLE cases and 27,911 controls). First, we evaluated the selected 8 SNPs in the replication study 1. In the replication study 2, 2 SNPs that satisfied $P<1.0\times10^{-6}$ in the combined study of GWAS and replication

study 1 were further evaluated (Figure 1). Among the evaluated SNPs, we observed significant replications in the SNP located in the genomic region of the AF4/FMR2 family, member 1 gene (*AFF1*) at 4q21 (rs340630; $P=4.6\times10^{-5}$ and $P=0.0094$ in the two individual cohorts, respectively; Table 3, Table S5, and Figure 2B). The combined study for the GWAS ($P=1.5\times10^{-4}$) and the replication studies demonstrated significant associations of rs340630 that satisfied the genome-wide significance threshold ($P=8.3\times10^{-9}$, OR = 1.21, 95% CI 1.14–2.30).

Cis-eQTL effect of rs340630 on *AFF1* transcripts

Since the landmark SNP in the *AFF1* locus, rs340630, was prioritized through the eQTL study as an eQTL positive SNP (Table 3), we further validated its cis-eQTL effect using Epstein-Barr virus (EBV)-transfected B cell lines established from Japanese individuals (Pharma SNP Consortium (PSC) cells, $n=62$). The correlation between rs340630 genotypes and the expression levels of *AFF1* was significant in the PSC cells stimulated with phorbol myristate acetate (PMA) ($R^2 = 0.074$, $P=0.033$; Figure 3A). The expression levels increased with the number of SLE-risk (A) alleles. To further confirm this cis-regulatory effect, we performed allele-specific transcript quantification (ASTQ) of *AFF1*. The transcript levels of each allele were quantified by qPCR using an allele specific probe for a SNP in the 5′-untranslated region (rs340638), which was in absolute LD with rs340630 ($r^2 = 1.0$, $D' = 1.0$). We examined PSC-cells ($n=17$) that were heterozygous for both rs340630 and rs340638. The mean ratio of each transcript (A over G allele; the A allele comprises a haplotype with the risk (A) allele of rs340630) were significantly increased to 1.07 compared to the ratio of the amount of DNA (1.00, $P=0.012$) (Figure 3B). These results suggest that rs340630, or SNP(s) in LD with it, are a regulatory variant predisposing SLE susceptibility through increased expression levels of *AFF1*.

Expression of *AFF1* in CD4⁺ and CD19⁺ peripheral blood lymphocytes

AFF1 is known to be involved in cytogenetic translocations of acute lymphoblastic leukemia (ALL) [31]. Its fusion protein with the mixed-lineage leukemia gene (*MLL*) is implicated in the regulation of transcription and the cell cycle of lymphocytes [31]. To investigate the expression pattern of *AFF1* in normal tissues, we evaluated the transcript levels of *AFF1* in a panel of various tissues. We observed prominent expression of *AFF1* in CD4⁺ and CD19⁺ peripheral blood lymphocytes, implying an important role for *AFF1* in helper-T-cells and B-cells (Figure 3C).

Discussion

Through a GWAS and multi-staged replication studies consisting of 2,278 SLE cases and 31,948 controls in Japanese subjects, our study identified that the *AFF1* locus was significantly associated with SLE susceptibility.

As well as the identification of the novel SLE susceptibility locus, we observed significant replications of associations in the previously reported susceptibility loci. The replications were especially enriched in the loci identified through the studies in Asian populations, compared to those in European populations. Considering the ethnical heterogeneities in the epidemiology of SLE [19,20], these observations suggest the similarities in the genetic backgrounds of SLE shared within Asian populations, and also the existence of the both common and divergent genetic backgrounds encompassed between European and Asian populations.

206

Table 1. Results of a genome-wide association study for Japanese patients with SLE.

rsID[a]	Chr	Position (bp)	Cytoband	Gene	Allele[b] 1/2	No. subjects Case	No. subjects Control	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P
rs10168266	2	191,644,049	2q32	STAT4	T/C	891	3,384	0.37	0.27	1.59 (1.42–1.78)	2.7×10^{-16}
rs9501626	6	32,508,322	6p21	HLA region	A/C	891	3,381	0.20	0.12	1.86 (1.62–2.13)	1.0×10^{-18}
rs2230926	6	138,237,759	6q23	TNFAIP3	G/T	891	3,377	0.11	0.069	1.75 (1.47–2.08)	1.9×10^{-10}
rs6964720	7	75,018,280	7q11	HIP1	G/A	891	3,384	0.25	0.19	1.43 (1.27–1.63)	1.3×10^{-8}
rs2254546	8	11,381,089	8p23	BLK	G/A	891	3,384	0.78	0.72	1.42 (1.61–1.25)	4.1×10^{-8}
rs6590330	11	127,816,269	11q24	ETS1	A/G	891	3,368	0.48	0.39	1.44 (1.30–1.60)	1.3×10^{-11}

[a]SNPs that satisfied the threshold of $P < 5.0 \times 10^{-8}$ were indicated.
[b]Based on forward strand of NCBI Build 36.3.
SLE, systemic lupus erythematosus; OR, odds ratio.
doi:10.1371/journal.pgen.1002455.t001

To effectively detect the novel SLE susceptibility locus, we integrated cis-eQTL effects of the SNPs and prioritized the results of the GWAS. In addition to identifying a novel locus for SLE-susceptibility, our study demonstrated approximately 30% of confirmed SLE-susceptibility loci were comprised of cis-eQTLs. We also confirmed cis-regulatory effect of the landmark SNP in the AFF1 locus, rs340630, on AFF1 transcripts, which had been prioritized through the eQTL study. These results would suggest that accumulation of quantitative changes in gene expression would accelerate the disease onset of SLE. It would also demonstrate the validity of applying eQTL study in the search of the susceptible genes for SLE or other autoimmune diseases, as previously suggested in the study for celiac disease [24]. To our knowledge, this is one of the initial studies to successfully discover a new locus by prioritizing GWAS results using eQTLs, and should contribute to the approaches assessing genetic loci still being uncaptured by recent large-scaled GWASs due to stringent significance threshold for multiple hypothesis testing [21].

We observed prominent expression levels of AFF1 in CD4+ and CD19+ peripheral blood lymphocytes, which would imply an important role for AFF1 in helper-T-cells and B-cells. In fact, AFF1 is essential for normal lymphocyte development, as demonstrated in mice deficient for AFF1; severe reduction were observed in the thymic double positive CD4/CD8 population and the bone marrow pre-B and mature B-cell numbers [32]. The risk A allele of rs340630 demonstrated a cis-eQTL effect on the AFF1 transcript with enhanced expression levels. As the AFF1 locus was also demonstrated as an eQTL in primary liver cells [33], the cis-regulatory effect may hold in primary cells as well as lymphoblastoid cells used in the present study. However, because the mechanism of transcriptional regulation is substantially different among cell types [34], cell-type specific analyses including those for primary T-cells and B-cells are needed for understanding the precise role of AFF1 variant in primary lymphocytes. Although further functional investigation is necessary, our observation suggested that AFF1 is involved in the etiology of SLE through the regulation of development and activity of lymphocytes. It is of note that AFF3, which also belongs to the AF4/FMR2 family, is associated with susceptibility to autoimmune diseases [35].

One of our study's limitations is the selection of SNPs for the replication study using the results of the pooled DNA approach [30], which used a different genotyping platform from that of the present GWAS. Moreover, the association signals based on Silhouette scores in pooled analysis would be less reliable compared to those based on individual genotyping. Since direct comparisons of the association signals of the same single SNPs

between the studies would be difficult due to these issues, we adopted the complementary approach that referred the association signals of the multiple SNPs in the pooled analysis for each of the single SNPs in the GWAS, taking account of LD and physical distances between the SNPs. However, there would exist a possibility that the variant(s) truly associated with SLE was left not to be examined in the replication study. It should be noted that only 1 SNP among the 8 selected SNPs yielded the significant association with SLE, although further enrichments of the significant associations might be anticipated. To elucidate effectiveness and limitation of our approach, further assessments of the studies on the remaining loci would be desirable. It should also be noted that the control-case ratio of the subjects were relatively high in the replication study 2 (= 33.8), and this disproportionate ratio could have induced potential bias on the results of the association analysis of the SNPs. However, considering the homogeneous ancestries of the Japanese population [27] and that principal component analysis did not demonstrate significant population stratification in the control subjects of the replication study 2 (data not shown), the bias owing to population stratification might not be substantial.

In summary, through a GWAS and multi-staged replication studies in a Japanese population integrating eQTL study, our study identified AFF1 as a novel susceptibility locus for SLE.

Materials and Methods

Subjects

We enrolled 2,278 systemic lupus erythematosus (SLE) cases and 31,948 controls. SLE cases enrolled in the genome-wide association study (GWAS) ($n = 891$) or part of the 2nd replication study ($n = 83$) were collected from 12 medical institutes in Japan under the support of the autoimmune disease study group of Research in Intractable Diseases, Japanese Ministry of Health, Labor and Welfare: Hokkaido University Graduate School of Medicine, Tohoku University Graduate School of Medicine, the University of Tokyo, Keio University School of Medicine, Juntendo University School of Medicine, University of Occupational and Environmental Health, University of Tsukuba, Tokyo Medical and Dental University, National Center for Global Health and Medicine, Nagasaki University, Wakayama Medical University, and Jichi Medical University. SLE cases ($n = 562$) and controls ($n = 653$) enrolled in the 1st replication study were collected from Kyushu University. Some of the SLE cases ($n = 742$) and controls ($n = 27,911$) enrolled in the 2nd replication study were collected from Kyoto University, Tokyo Women's
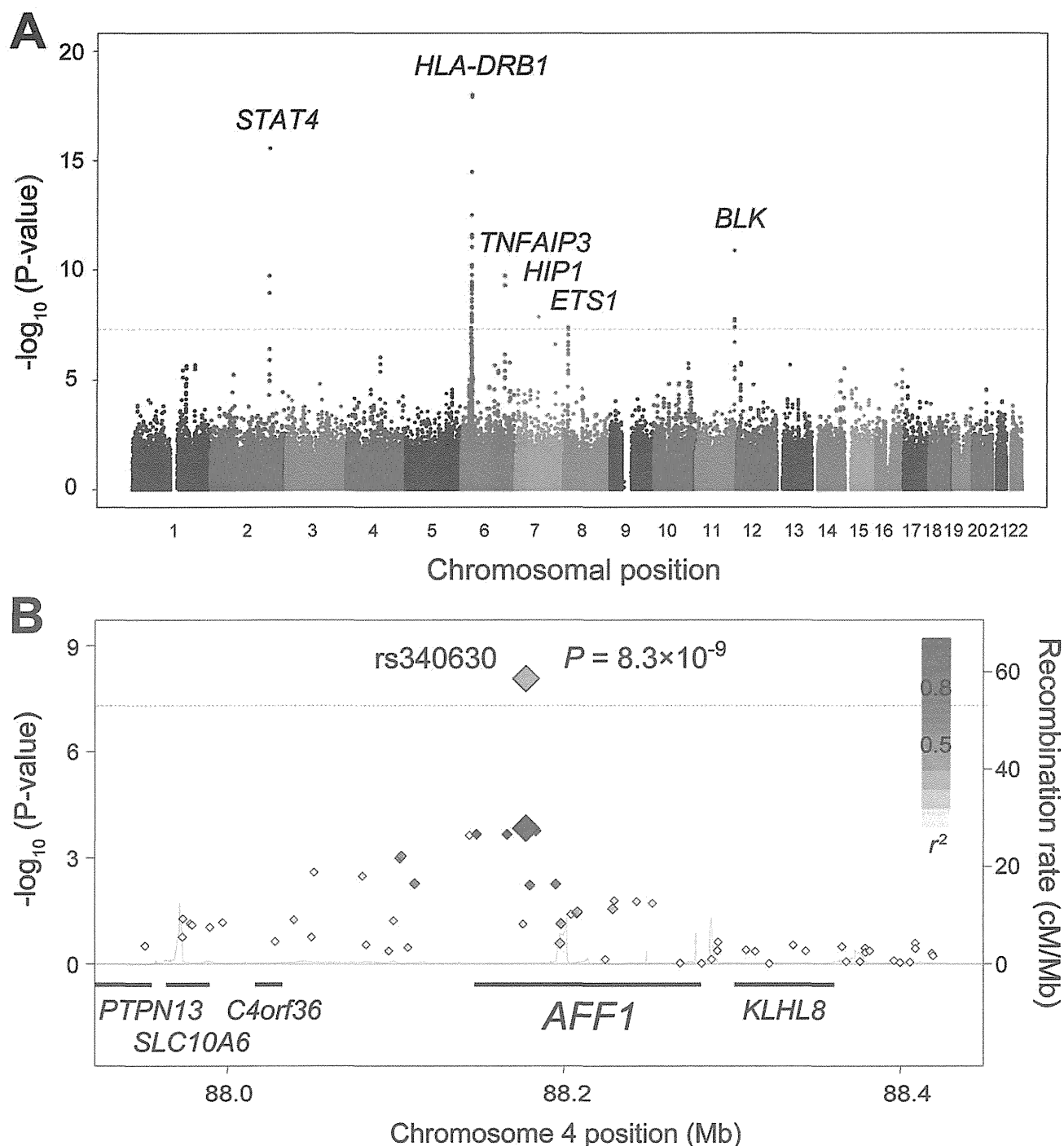
Figure 2. Associations of the *AFF1* locus with SLE. (A) A chromosomal plot of P-values in GWAS for SLE. (B) A regional plot in the *AFF1* locus. Diamond-shaped data points represent $-\log_{10}$ (P-values) of the SNPs. Large-sized points indicate the P-values of the landmark SNP, rs340630 (green for the combined study and red for the GWAS). Density of red color represents r^2 values with rs340630. Blue line represents recombination rates. Lower part indicates RefSeq genes. Gray dashed horizontal lines represent the threshold of $P=5.0\times10^{-8}$. The plots were drawn using SNAP, version 2.1 [47].
doi:10.1371/journal.pgen.1002455.g002

Medical University, the University of Tokyo, and the BioBank Japan Project [36]. All subjects were of Japanese origin and provided written informed consent. SLE cases met the revised American College of Rheumatology (ACR) criteria for SLE [37]. Control subjects were confirmed to be free of autoimmune disease. Some of the SLE cases were included in our previous studies [38–40]. Details of the subjects are summarized in Table S1 and S2. This research project was approved by the ethical committees of the University of Tokyo, RIKEN, and affiliated medical institutes.

208

Table 2. Associations among previously reported SLE-related loci.

rsID	Chr	Position (bp)	Cytoband	Gene	Allele[a] 1/2	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P	eQTL[b]	Identified by the studies in[c] Caucasians	Identified by the studies in[c] Asians
rs2205960	1	171,458,098	1q25	TNFSF4	T/G	0.23	0.18	1.35 (1.19–1.54)	3.0×10^{-6}		+	
rs3024505	1	205,006,527	1q32	IL10	A/G	0.019	0.014	1.34 (0.90–2.00)	0.15		+	
rs13385731	2	33,555,394	2p22	RASGRP3	C/T	0.90	0.87	1.37 (1.15–1.64)	6.0×10^{-4}	+		+
rs10168266	2	191,644,049	2q32	STAT4	T/C	0.37	0.27	1.59 (1.42–1.78)	2.7×10^{-16}		+	
rs6445975	3	58,345,217	3p14	PXK	G/T	0.25	0.23	1.09 (0.96–1.23)	0.18	+	+	
rs10516487	4	102,970,099	4q24	BANK1	G/A	0.91	0.89	1.28 (1.07–1.53)	0.0070		+	
rs10036748	5	150,438,339	5q33	TNIP1	T/C	0.75	0.72	1.16 (1.03–1.31)	0.014			+
rs9501626	6	32,508,322	6p21	HLA-DRB1	A/C	0.20	0.12	1.86 (1.62–2.13)	1.0×10^{-18}		+	
rs548234	6	106,674,727	6q21	PRDM1	C/T	0.40	0.34	1.30 (1.16–1.44)	2.3×10^{-6}	+	+	
rs2230926	6	138,237,759	6q23	TNFAIP3	G/T	0.11	0.069	1.75 (1.47–2.08)	1.9×10^{-10}	+	+	
rs849142	7	28,152,416	7p15	JAZF1	C/T	0.999	0.999	2.72 (0.25–29.8)	0.41		+	
rs4917014	7	50,276,409	7p12	IKZF1	T/G	0.58	0.53	1.24 (1.11–1.38)	8.1×10^{-5}			+
rs6964720	7	75,018,280	7q11	HIP1	G/A	0.25	0.19	1.43 (1.27–1.62)	1.3×10^{-8}			+
rs4728142	7	128,361,203	7q32	IRF5	A/G	0.16	0.11	1.48 (1.28–1.72)	2.4×10^{-7}	+	+	
rs2254546	8	11,381,089	8p23	BLK	G/A	0.78	0.72	1.42 (1.25–1.61)	4.1×10^{-8}	+	+	
rs1913517	10	49,789,060	10q11	WDFY4	A/G	0.32	0.28	1.20 (1.07–1.35)	0.0013			+
rs4963128	11	579,564	11p15	KIAA1542	T/C	0.98	0.97	1.58 (1.03–2.44)	0.038	+	+	
rs2732552	11	35,041,168	11p13	PDHX, CD44	T/C	0.75	0.73	1.13 (1.00–1.27)	0.056		+	
rs4639966	11	118,078,729	11q23	Intergenic	T/C	0.32	0.28	1.22 (1.09–1.36)	7.3×10^{-4}			+
rs6590330	11	127,816,269	11q24	ETS1	A/G	0.48	0.39	1.44 (1.30–1.60)	1.3×10^{-11}			+
rs1385374	12	127,866,647	12q24	SLC15A4	T/C	0.19	0.16	1.21 (1.06–1.38)	0.0057			+
rs7329174	13	40,456,110	13q14	ELF1	G/A	0.30	0.25	1.32 (1.18–1.49)	2.2×10^{-6}			+
rs7197475	16	30,550,368	16p11	Intergenic	T/C	0.12	0.10	1.20 (1.02–0.41)	0.031			+
rs11150610	16	31,241,737	16p11	ITGAM	C/A	0.20	0.19	1.07 (0.94–1.22)	0.32	+	+	
rs12949531	17	13,674,531	17p12	Intergenic	T/C	0.28	0.27	1.02 (0.91–1.15)	0.73		+	
rs463426	22	20,139,185	22q11	HIC2,UBE2L3	T/C	0.52	0.48	1.20 (1.08–1.33)	6.1×10^{-4}		+	

[a]Based on forward strand of NCBI Build 36.3.
[b]Defined using gene expression data measured in lymphoblastoid B cell lines [28].
[c]Based on the previously reported studies for SLE susceptibility loci [3–18].
SLE, systemic lupus erythematosus; OR, odds ratio; eQTL, expression quantitative trait locus; GWAS, genome-wide association study.
doi:10.1371/journal.pgen.1002455.t002

Genotyping and quality control

In GWAS, 946 SLE cases and 3,477 controls were genotyped using Illumina HumanHap610-Quad and Illumina Human- Hap550v3 Genotyping BeadChips (Illumina, CA, USA), respectively. After the exclusion of 47 SLE cases and 92 controls with call rates <0.98, SNPs with call rates <0.99 in SLE cases or controls,

Table 3. Results of combined study for Japanese patients with SLE.

rsID	Chr	Position (bp)	Cytoband	Gene	Allele 1/2	Stage	No. subjects Case	No. subjects Control	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P	eQTL[a]
rs340630	4	88,177,419	4q21	AFF1	A/G	GWAS	891	3,383	0.56	0.51	1.22 (1.10–1.36)	1.5×10^{-4}	+
						Replication study 1	550	646	0.57	0.49	1.40 (1.19–1.64)	4.6×10^{-5}	
						Replication study 2	820	27,911	0.56	0.53	1.14 (1.03–1.26)	0.0094	
						Combined study	2,261	31,940	0.56	0.52	1.21 (1.14–1.30)	8.3×10^{-9}	

[a]Defined using gene expression data measured in lymphoblastoid B cell lines [28].
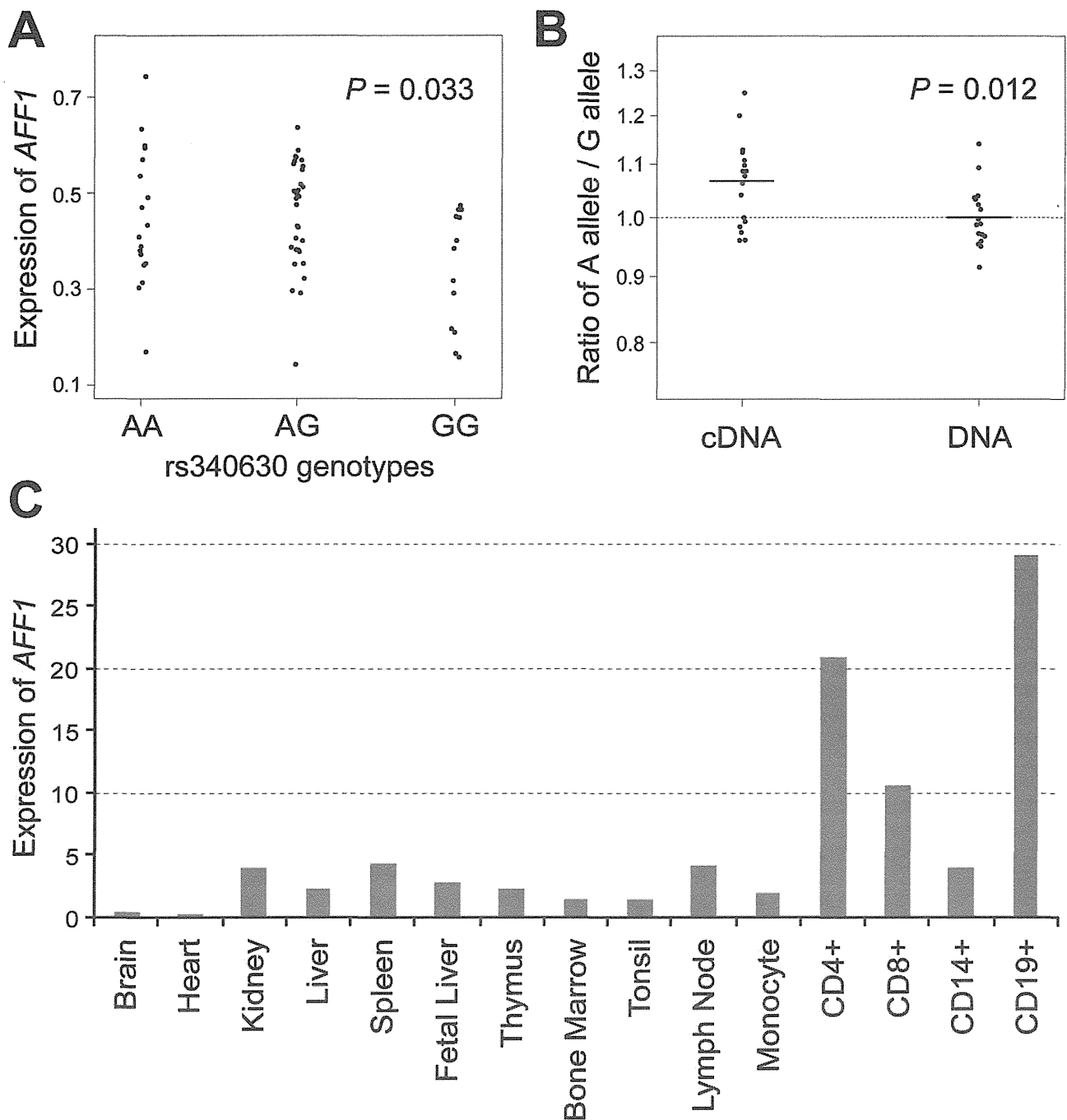doi:10.1371/journal.pgen.1002455.t003

Figure 3. Association of rs340630 with *AFF1* expression. (A) Correlation between rs340630 genotypes and transcript levels of *AFF1* (NM_001166693) in EBV-transfected cell lines ($n = 62$) stimulated with PMA. (B) Allele-specific quantification (ASTQ) of *AFF1* transcripts. Allele specific-probes for rs340638 were used for quantification by qPCR. The ratios of A allele over G allele for the amounts of both cDNAs and DNAs were plotted in log scale for each cell line. (C) *AFF1* expression in various tissues. Transcripts levels of *AFF*1 were quantified by qPCR and were normalized by *GAPDH* levels.
doi:10.1371/journal.pgen.1002455.g003

non-autosomal SNPs, and SNPs not shared between SLE cases and controls, were excluded. We excluded 7 closely related SLE cases in a 1st or 2nd degree of kinship based on identity-by-descent estimated using PLINK version 1.06 [41]. We then excluded 1 SLE cases and 1 controls whose ancestries were estimated to be distinct from East-Asian populations using PCA performed along with the genotype data of Phase II HapMap populations (release 24) [29] using EIGENSTRAT version 2.0 [42]. Subsequently,

SNPs with minor allele frequencies <0.01 in SLE cases or controls, SNPs with exact P-values of Hardy-Weinberg equilibrium test $<1.0 \times 10^{-6}$ in controls, or SNPs with ambiguous cluster plots were excluded. Finally, 430,797 SNPs for 891 SLE cases and 3,384 controls were obtained. Genotyping of SNPs in replication studies was performed using TaqMan Assay or Illumina HumanHap610-Quad Genotyping BeadChip (Illumina, CA, USA).

Association analysis of the SNPs

Association of SNPs in GWAS and replication studies were tested with Cochran-Armitage's trend test. Combined analysis was performed with Mantel-Haenzel method. Associations of previously reported SLE susceptibility loci [3–18] were evaluated using the results of the GWAS. Genotype imputation was performed for non-genotyped SNPs using MACH version 1.0 [43] with Phase II HapMap East-Asian individuals as references [29], as previously described [44]. All imputed SNPs demonstrated imputation scores, Rsq, >0.70.

eQTL study

We analyzed gene expression data previously measured in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals using Illumina's human whole-genome expression array (WG-6 version 1) (accession number; GSE6536) [28]. Expression data were normalized across the individuals. We used BLAST to map 47,294 Illumina array probes onto human autosomal reference genome sequences (Build 36). We discarded probes mapped with expectation values smaller than 0.01 to multiple loci, or for which there was polymorphic HapMap SNP(s) inside the probe. Then, 19,047 probes with exact matches to a unique locus with 100% identity and with a mean signal intensity greater than background were obtained. Genotype data of HapMap individuals were obtained for SNPs included in the GWAS. Associations of SNP genotypes (coded as 0, 1, and 2) with expression levels of each of the cis-eQTL probes (located within ±300 kbp regions of the SNPs) were evaluated using linear regression assuming additive effects of the genotypes on the expression levels. Considering the significant overlap between eQTL and genetic loci responsible for autoimmune diseases [24], we applied relatively less stringent multiple testing threshold of FDR Q-values<0.2 for the definition of eQTL. SNPs that exhibited this threshold with any of the corresponding cis-eQTL probes were denoted as eQTL positive.

Selection of SNPs enrolled in the replication studies

In order to select SNPs for further replication studies, we firstly integrated the results of GWAS and eQTL study. SNPs that satisfied $P<1.0\times10^{-4}$ in GWAS, or the SNPs that satisfied $1.0\times10^{-4}\leq P<1.0\times10^{-3}$ in GWAS and denoted as eQTL positive, were selected. Among these, SNPs most significantly associated in each of the genomic loci and not included in the previously reported SLE susceptibility loci [3–18] were further evaluated.

Then, the results of the concurrently proceeding genome-wide scan for SLE in the Japanese subjects using a pooled DNA approach were referred (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009). In the scan, DNA collected from 447 SLE cases and 680 controls of Japanese origin were pooled respectively, and genotyped using GeneChip Human Mapping 500K Array Set (Affymetrix, CA, USA). SNPs were ranked according to the Silhouette scores estimated based on relative allele scores (RAS) between SLE cases and controls, and rank-based P-values were assigned [30]. By referring to association signals in multiple neighboring SNPs in the pooled analysis, we selected SNPs for replication study 1. Namely, if the SNP of interest was in LD ($r^2>0.5$) or was located within ±100 kbp of SNPs showing association signals in the pooled analysis (rank-based $P<0.01$), it would be selected. SNPs that satisfied $P<1.0\times10^{-6}$ in the combined study of GWAS and replication study 1 were further evaluated in replication study 2 (Figure 1).

Quantification of *AFF1* expression

EBV-transformed lymphoblastoid cell lines ($n=62$) were established by Pharma SNP Consortium (Tokyo, Japan) using peripheral blood lymphocytes of Japanese healthy individuals. Cells were incubated for 2 h in medium alone (RPMI 1640 medium containing 10% FBS, 1% penicillin, and 1% streptomycin) or with 100 ng/ml PMA. Conditions for cell stimulation were optimized before the experiment as previously described [45]. Cells were then harvested and total RNA was isolated using an RNeasy Mini Kit (Qiagen) with DNase treatment. Total RNA (1 μg) was reverse transcribed using TaqMan Gold RT-PCR reagents with random hexamers (Applied Biosystems). Real-time quantitative PCR was performed in triplicate using an ABI PRISM 7900 and TaqMan gene expression assays (Applied Biosystems). Specific probes (Hs01089428_m1) for transcript of *AFF1* (NM_001166693) were used. Expression of *AFF1* in various tissues was also quantified using Premium Total RNA (Clontech). The data were normalized to *GAPDH* levels. *GUS* levels were also evaluated for internal control, and similar results were obtained. Correlation coefficient, R^2, between rs340630 genotypes and transcript levels of *AFF1* was evaluated.

Allele-specific transcript quantification (ASTQ)

ASTQ of *AFF1* in PSC cells was performed as previously described [46]. DNAs were extracted by using a DNeasy Kit (QIAGEN). RNA extraction and cDNA preparation were performed as described above. For PSC cells ($n=17$) that were heterozygous for both rs340630 (the landmark SNP of GWAS) and rs340638 (located in the 5'-untranslated region of *AFF1* and in absolute LD with rs340630), expression levels of *AFF1* were quantified by qPCR on an ABI Prism 7900 using a custom-made TaqMan MGB-probe set for rs340638. Primer sequences were 5'-CTAACTGTGGCCCGCGTTG-3' and 5'-CCCGGCGCA-GTTTCTGAG-3'. The probe sequences were 5'-VIC-CGAA-GACCGCCAGCGCCCAAC-TAMRA-3' and 5'-FAM-CGAA-GACCGCCGGCGCCCAA-TAMRA-3'. Ct values of VIC and FAM were obtained for genomic DNA and cDNA samples after 40 cycles of real-time PCR. We also prepared genomic DNA of samples homozygous for each allele and mixed them at different ratios (2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2) to create a standard curve by plotting Ct values of VIC/FAM against the allelic ratio of VIC/FAM for each mixture. Using the standard curve, we calculated the allelic ratios for each genomic DNA and cDNA samples. We measured each sample in quadruplicate in one assay; tests were independently repeated twice.

Web resources

The URLs for data presented herein are as follows.
NCBI GEO, http://www.ncbi.nlm.nih.gov/geo
BioBank Japan Project, http://biobankjp.org
PLINK software, http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml
International HapMap Project, http://www.hapmap.org
EIGENSTRAT software, http://genepath.med.harvard.edu/~reich/Software.htm
MACH and mach2qtl software, http://www.sph.umich.edu/csg/abecasis/MACH/index.html
SNAP, http://www.broadinstitute.org/mpg/snap/index.php

Supporting Information

Figure S1 Principal component analysis (PCA) plot of the subjects. PCA plot of subjects enrolled in the GWAS for SLE. SLE cases and the controls enrolled in the GWAS are plotted based on

211

eigenvectors 1 and 2 obtained from the PCA using EIGEN-STRAT version 2.0 [42], along with European (CEU), African (YRI), Japanese (JPT), and Chinese (CHB) individuals obtained from the Phase II HapMap database (release 22) [29]. Subjects who were estimated to be outliers in terms of ancestry from East-Asian (JPT+CHB) clusters and excluded from the study are indicated by black arrows.
(TIF)

Figure S2 Quantile-Quantile plot (QQ-plot) of *P*-values in the GWAS for SLE. The horizontal axis indicates the expected $-\log_{10}$ (*P*-values). The vertical axis indicates the observed $-\log_{10}$ (*P*-values). The QQ-plot for the *P*-values of all SNPs that passed the quality control criteria is indicated in blue. The QQ-plot for the *P*-values after the removal of SNPs included in the previously reported SLE susceptibility loci is indicated in black. The gray line represents $y = x$. The SNPs for which the *P*-value was smaller than 1.0×10^{-15} are indicated at the upper limit of the plot.
(TIF)

Table S1 Basal characteristics of cohorts.
(DOC)

Table S2 Frequency of clinical characteristics of SLE in this GWAS.
(DOC)

Table S3 Distributions of eQTL positivity rates of the SNPs.
(DOC)

Table S4 Results of replication study 1 for Japanese patients with SLE.
(DOC)

Table S5 Results of replication studies 1 and 2 for Japanese patients with SLE.
(DOC)

Author Contributions

Conceived and designed the experiments: Y. Okada, K. Shimane, Y. Kochi, A. Suzuki, M. Kubo, N. Kamatani, R. Yamada, Y. Nakamura, K. Yamamoto. Performed the experiments: Y. Kochi, N. Hosono, M. Kubo, K. Myouzen, T. Horiuchi, T. Tahira, K. Hayashi. Analyzed the data: Y. Okada, A. Takahashi, T. Tsunoda, K. Higasa, R. Yamada. Contributed reagents/materials/analysis tools: T. Horita, T. Atsumi, T. Koike, T. Ishii, A. Okamoto, K. Fujio, M. Hirakata, H. Amano, Y. Takasaki, Y. Kondo, S. Ito, T. Sumida, K. Takada, A. Mimori, K. Saito, Y. Tanaka, M. Kamachi, N. Nishimoto, Y. Kawaguchi, C. Terao, K. Ohmura, T. Mimori, F. Matsuda, O.W. Mohammed, K. Matsuda, K. Shimane, K. Ikari, H. Yamanaka. Wrote the paper: Y. Okada, K. Shimane, Y. Kochi, A. Suzuki, R. Yamada, K. Yamamoto.

References

1. Lipsky PE (2001) Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. Nat Immunol 2: 764–766.
2. Sestak AL, Shaver TS, Moser KL, Neas BR, Harley JB (1999) Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. J Rheumatol 26: 1495–1499.
3. Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, et al. (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet 76: 528–537.
4. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, et al. (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 38: 550–555.
5. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, et al. (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci U S A 104: 6758–6763.
6. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357: 977–986.
7. Cunninghame Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, et al. (2008) Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. Nat Genet 40: 83–89.
8. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, et al. (2008) A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet 40: 152–154.
9. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, et al. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40: 204–210.
10. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, et al. (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40: 211–216.
11. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, et al. (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 358: 900–909.
12. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, et al. (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 40: 1059–1061.
13. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, et al. (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet 40: 1062–1064.
14. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, et al. (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 41: 1234–1237.
15. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, et al. (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet 41: 1228–1233.
16. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, et al. (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet 6: e1000841. doi:10.1371/journal.pgen.1000841.
17. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, et al. (2011) Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. Am J Hum Genet 88: 83–91.
18. Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, et al. (2011) ELF1 is associated with systemic lupus erythematosus in Asian populations. Hum Mol Genet 20: 601–607.
19. Hopkinson ND, Doherty M, Powell RJ (1994) Clinical features and race-specific incidence/prevalence rates of systemic lupus erythematosus in a geographically complete cohort of patients. Ann Rheum Dis 53: 675–680.
20. Danchenko N, Satia JA, Anthony MS (2006) Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus 15: 308–318.
21. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569.
22. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, et al. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet 5: e1000534. doi:10.1371/journal.pgen.1000534.
23. Cantor RM, Lange K, Sinsheimer JS (2010) Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet 86: 6–22.
24. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302.
25. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184–194.
26. Kochi Y, Okada Y, Suzuki A, Ikari K, Terao C, et al. (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet 42: 515–519.
27. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, et al. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83: 445–456.
28. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224.
29. The International HapMap Consortium (2003) The International HapMap Project. Nature 426: 789–796.

212

30. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, et al. (2007) Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet 80: 126–139.
31. Xia ZB, Popovic R, Chen J, Theisler C, Stuart T, et al. (2005) The MLL fusion gene, MLL-AF4, regulates cyclin-dependent kinase inhibitor CDKN1B (p27kip1) expression. Proc Natl Acad Sci U S A 102: 14028–14033.
32. Isnard P, Core N, Naquet P, Djabali M (2000) Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. Blood 96: 705–710.
33. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107. doi:10.1371/journal.pbio.0060107.
34. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43–49.
35. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514.
36. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.
37. Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40: 1725.
38. Suzuki A, Yamada R, Kochi Y, Sawada T, Okada Y, et al. (2008) Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. Nat Genet 40: 1224–1229.
39. Shimane K, Kochi Y, Horita T, Ikari K, Amano H, et al. (2010) The association of a nonsynonymous single-nucleotide polymorphism in TNFAIP3 with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population. Arthritis Rheum 62: 574–579.
40. Myouzen K, Kochi Y, Shimane K, Fujio K, Okamura T, et al. (2010) Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus. Hum Mol Genet 19: 2313–2320.
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
42. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.
43. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10: 387–406.
44. Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, et al. (2011) Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum Mol Genet 20: 1224–1231.
45. Aikawa Y, Yamamoto M, Yamamoto T, Morimoto K, Tanaka K (2002) An anti-rheumatic agent T-614 inhibits NF-kappaB activation in LPS- and TNF-alpha-stimulated THP-1 cells without interfering with IkappaBalpha degradation. Inflamm Res 51: 188–194.
46. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, et al. (2010) A functional variant in NKX3.1 associated with prostate cancer susceptibility down-regulates NKX3.1 expression. Hum Mol Genet 19: 4265–4272.
47. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24: 2938–2939.

213

Letter

Quantitative effect of HLA-DRB1 alleles to ACPA levels in Japanese rheumatoid arthritis: no strong genetic impact of shared epitope to ACPA levels after stratification of HLA-DRB1*09:01

Anti-citrullinated peptide antibody (ACPA) is a highly specific serological marker for rheumatoid arthritis (RA).[1-3] Different HLA-DRB1 alleles have been shown to be associated with the susceptibility to ACPA-positive RA.[4][5] Former studies demonstrated that HLA-DRB alleles carrying a shared epitope (SE),[6] consisting of a conserved amino acid motif at positions 70–74 of the HLA-DRβ chain, were strongly associated with ACPA-positive RA and with higher ACPA levels in European and Japanese populations.[7-9] On the other hand, HLA-DRB1*09:01 was recently found to be negatively associated with ACPA levels in the Japanese.[9] These observations imply that combinations of HLA-DRB1 alleles differentially influence ACPA levels in ACPA-positive RA.

To address this question, we conducted a genetic association study employing 2457 ACPA-positive Japanese RA patients. ACPA was quantified by MESACUP CCP ELISA kit (MBL Co Ltd, Nagoya, Japan) with a cut-off level of 4.5 U/ml. The patients were then divided into three groups based on their ACPA titres: low (~4.5–13.5 U/ml), intermediate (~13.5–100 U/ml) and high (≥100 U/ml) groups. These groups were defined according to the 2010 ACR/EULAR classification criteria for RA and a measurement limit of the kit. HLA-DRB1 genotyping was carried out using either the Wakflow system (Wakunaga Pharmaceutical Co Ltd, Osaka, Japan) or the sequencing-based AlleleSEQR HLA-DRB1 typing kit (Abbott Japan, Nagoya, Japan). Frequencies of HLA-DRB1 alleles were compared among the three groups using the Cochran-Armitage Trend test. The relative predispositional effect (RPE) method was applied to identify the associations of more than one HLA-DBR1 allele sequentially according to their strength.[10] Briefly, associations of HLA-DRB1 alleles with ACPA categories were estimated for each allele using the Cochran-Armitage Trend test. When we detected the strongest association with a significant p value, the allele was excluded from the whole data and the same steps were repeated until no further significant alleles were found.

As expected from the previous studies,[9] HLA-DRB1*09:01 showed the strongest association with ACPA levels in a decreasing manner (p=1.0×10^{-21}) and the SE alleles were significantly associated with an increasing effect (p=3.2×10^{-7}) (table 1). In addition, HLA-DRB1*04:07 showed negative association with ACPA levels (p=0.0013), and HLA-DRB1*15:01 and HLA-DRB1*15:02 were positively associated with ACPA levels (p=2.3×10^{-5} and 0.0011, respectively) (table 1). Of note, the association between the SE and ACPA levels lost significance after

Table 1 Association of HLA-DRB1 alleles with ACPA levels

HLA-DRB1	Low n=594	Intermediate n=1510	High n=2810	p Value	RPE p Value	RPE (OR)	Effect on ACPA levels
SE							
SEall	216 (36.4%)	616 (40.8%)	1303 (46.4%)	3.2×10^{-7}	0.16†	1.08 (0.98–1.20)†	
DRB1*01:01	32 (5.4%)	96 (6.4%)	223 (7.9%)	0.0096			
DRB1*04:01	18 (3.0%)	47 (3.1%)	82 (2.9%)	0.78			
DRB1*04:04	2 (0.3%)	1 (0.1%)	14 (0.5%)	0.13			
DRB1*04:05	138 (23.2%)	409 (27.1%)	840 (29.9%)	0.00053			
DRB1*04:10	17 (2.9%)	33 (2.2%)	67 (2.4%)	0.71			
DRB1*10:01	6 (1.0%)	13 (0.9%)	28 (1.0%)	0.87			
DRB1*14:06	3 (0.5%)	14 (0.9%)	44 (1.6%)	0.013			
Non-SE							
DRB1*04:03	12 (2.0%)	30 (2.0%)	31 (1.1%)	0.019			
DRB1*04:06	17 (2.9%)	14 (0.9%)	57 (2.0%)	0.96			
DRB1*04:07	5 (0.8%)	11 (0.7%)	4 (0.1%)	0.0013	0.00034	0.30 (0.16–0.57)	(−)
DRB1*08:02	15 (2.5%)	30 (2.0%)	60 (2.1%)	0.74			
DRB1*08:03	36 (6.1%)	66 (4.4%)	119 (4.2%)	0.10			
DRB1*09:01	158 (26.6%)	334 (22.1%)	367 (13.1%)	1.0×10^{-21}	1.0×10^{-21}	0.56 (0.50–0.62)	(−)
DRB1*11:01	8 (1.3%)	27 (1.8%)	50 (1.8%)	0.57			
DRB1*12:01	14 (2.4%)	30 (2.0%)	68 (2.4%)	0.63			
DRB1*12:02	8 (1.3%)	26 (1.7%)	50 (1.8%)	0.52			
DRB1*13:02	22 (3.7%)	53 (3.5%)	102 (3.6%)	0.98			
DRB1*14:01	4 (0.7%)	32 (2.1%)	32 (1.1%)	0.64			
DRB1*14:03	6 (1.0%)	17 (1.1%)	37 (1.3%)	0.46			
DRB1*14:05	5 (0.8%)	19 (1.3%)	21 (0.7%)	0.36			
DRB1*15:01	20 (3.4%)	53 (3.5%)	180 (6.4%)	2.3×10^{-5}	0.0011	1.53 (1.21–1.92)	(+)
DRB1*15:02	36 (6.1%)	120 (7.9%)	276 (9.8%)	0.0011			
DRB1*16:02	4 (0.7%)	20 (1.3%)	29 (1.0%)	0.83			

HLA-DRB1 alleles with frequencies greater than 0.5% are shown. Significant levels were set as 0.0022 for HLA-DRB1 alleles after Bonferroni's correction for multiple testing.
†p Value and OR after removal of HLA-DRB1*09:01.
ACPA, anti-citrullinated peptide antibody; RPE, relative predispositional effect; SE, shared epitope.
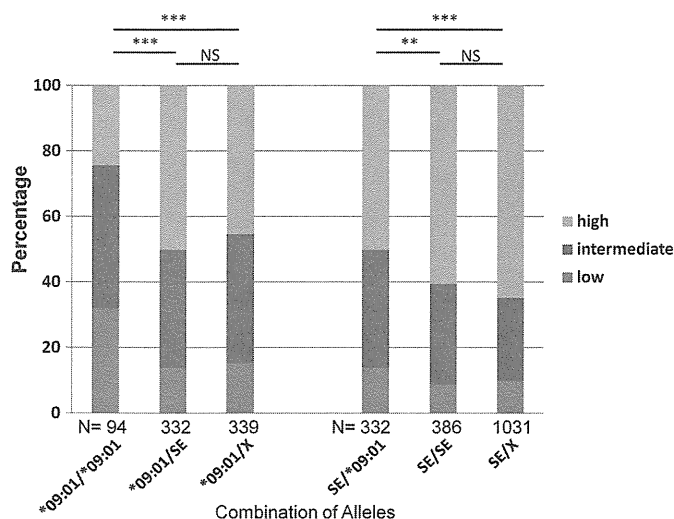
214

Letter

Figure 1 Comparisons of blood anti-citrullinated peptide antibody (ACPA) levels among HLA-DRB1*09:01, shared epitope (SE) and other alleles in combination. Frequencies of three rheumatoid arthritis subgroups based on ACPA levels were compared among different HLA-DRB1 combinations containing HLA-DRB1*09:01 and/or SE. X indicates HLA-DRB1 alleles other than HLA-DRB1*09:01 and SE. 'Low', 'intermediate' and 'high' categories correspond to patients with ACPA titres of ~4.5–13.5, ~13.5–100 and ≥100 U/ml, respectively. **p<0.005 and ***p<0.00005. NS, not significant.

stratification of HLA-DRB1*09:01 using RPE (p=0.16) whereas HLA-DRB1*04:07 and HLA-DRB1*15:01 remained significant after RPE (p=0.00034 and p=0.0011, respectively) (table 1). To confirm the dominant effect of HLA-DRB1*09:01 on ACPA levels over SE, we compared ACPA levels in two sets: first between HLA-DRB1*09:01/*09:01 and HLA-DRB1*09:01/SE or HLA-DRB1*09:01/X, and second between SE/HLA-DRB1*09:01 and SE/SE or SE/X. We found that HLA-DRB1*09:01 showed a significant association with low ACPA category compared with the other two groups in both sets of analyses (p<0.005, figure 1). On the other hand, we could not observe any difference between SE and the other alleles.

In this study, we aimed to identify HLA-DRB1 alleles showing quantitative effects on ACPA levels using a large collection of Japanese ACPA-positive RA patients. RPE was applied to avoid misleading frequency deviation by the allele with the strongest association to other associated alleles. We demonstrated that HLA-DRB1*09:01 was the strongest genetic determinant for lower ACPA levels, and the quantitative effects of HLA-DRB1 alleles carrying the SE were not a primary effect but merely an expected consequence of the decreased frequency of HLA-DRB1*09:01. We also identified two novel HLA-DRB1 alleles, HLA-DRB1*04:07 and HLA-DRB1*15:01, being associated with ACPA levels. It is interesting and feasible to perform similar studies in other populations and investigate whether or not the same set of HLA-DRB1 alleles are related to the quantitative effects beyond ethnicities and to examine if such alleles share conserved amino acid motifs.

Chikashi Terao,[1,2] Katsunori Ikari,[3] Koichiro Ohmura,[2] Taku Suzuki,[3] Takuji Iwamoto,[3] Kiyoshi Takasugi,[4] Hiroo Saji,[5] Atsuo Taniguchi,[3] Shigeki Momohara,[3] Hisashi Yamanaka,[3] Fumihiko Matsuda,[1,6,7] Tsuneyo Mimori[2]

[1]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[2]Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan
[3]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan
[4]Dohgo Spa Hospital, Matsuyama, Japan
[5]HLA Laboratory, Kyoto, Japan
[6]CREST Program, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan
[7]Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan

Correspondence to Chikashi Terao, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan; a0001101@kuhp.kyoto-u.ac.jp

REFERENCES

1. **van Venrooij WJ**, Hazes JM, Visser H. Anticitrullinated protein/peptide antibody and its role in the diagnosis and prognosis of early rheumatoid arthritis. *Neth J Med* 2002;**60**:383–8.
2. **Schellekens GA**, Visser H, de Jong BA, et al. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis Rheum* 2000;**43**:155–63.
3. **van Jaarsveld CH**, ter Borg EJ, Jacobs JW, et al. The prognostic value of the antiperinuclear factor, anti-citrullinated peptide antibodies and rheumatoid factor in early rheumatoid arthritis. *Clin Exp Rheumatol* 1999;**17**:689–97.
4. **van der Woude D**, Lie BA, Lundström E, et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. *Arthritis Rheum* 2010;**62**:1236–45.
5. **Gorman JD**, Lum RF, Chen JJ, et al. Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis Rheum* 2004;**50**:400–12.
6. **Gregersen PK**, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;**30**:1205–13.
7. **Balsa A**, Cabezón A, Orozco G, et al. Influence of HLA-DRB1 alleles in the susceptibility of rheumatoid arthritis and the regulation of antibodies against citrullinated proteins and rheumatoid factor. *Arthritis Res Ther* 2010;**12**:R62.
8. **Ohmura K**, Terao C, Maruya E, et al. Anti-citrullinated peptide antibody-negative RA is a genetically distinct subset: a definitive study using only bone-erosive ACPA-negative rheumatoid arthritis. *Rheumatology (Oxford)* 2010;**49**:2298–304.
9. **Okada Y**, Suzuki A, Yamada R, et al. HLA-DRB1*0901 lowers anti-cyclic citrullinated peptide antibody levels in Japanese patients with rheumatoid arthritis. *Ann Rheum Dis* 2010;**69**:1569–70.
10. **Payami H**, Joe S, Farid NR, et al. Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. *Am J Hum Genet* 1989;**45**:541–6.