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

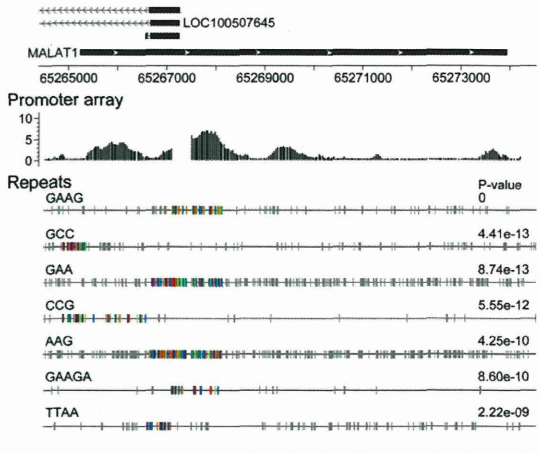

Fig. 4. Repeat sequences surrounding the breakage region in the *MALAT1* gene. (*Top*) Representation of a 10-kb segment surrounding the *MALAT1* locus. *x* axis numbers represent base positions according to hg19 assembly. (*Middle*) Breakage signal distribution detected by promoter array. Regions without 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.

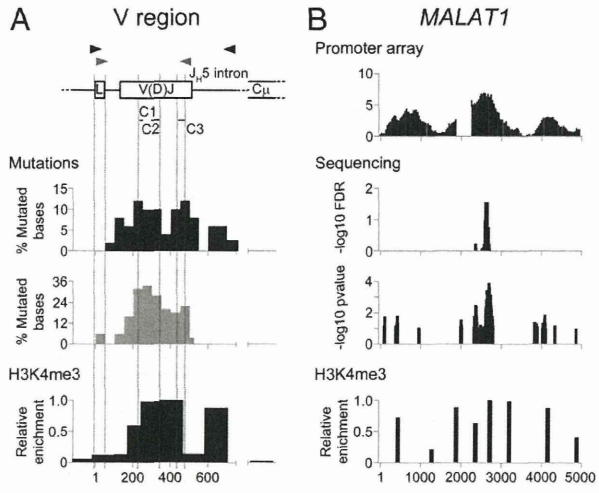

Fig. 5. H3K4me3 distribution in the IgH V region and in the *MALAT1* gene. (*A Top*: Representation of the rearranged IgH V region of BL2 cells. Black and gray arrowheads represent the position of primers used for the mutation analysis shown in *Bottom* (graphs in black and gray, respectively). L, leader; 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 5*B*). 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, TTTTTGTATTTT, 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*.

ACKNOWLEDGMENTS. We thank Y. Shiraki for manuscript preparation, Dr. H. Nagaoka for sharing unpublished BL2 expression data, and RIKEN Genome Network Analysis Service (GeNAS) for library sequencing using the

SOLiD system (Life Technologies). This research was supported by a RIKEN Omics Science Center from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan research grant (to Y.H.) and MEXT of Japan Grant-in-Aid for Specially Promoted Research 17002015.

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.

OPEN ACCESS Freely available online PLoS GENETICS

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.

Citation: Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, et al. (2012) A Genome-Wide Association Study Identified *AFF1* as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese. PLoS Genet 8(1): e1002455. doi:10.1371/journal.pgen.1002455

Editor: Mark I. McCarthy, University of Oxford, United Kingdom

Received May 16, 2011; **Accepted** November 18, 2011; **Published** January 26, 2012

Funding: This study was supported by a grant from the CGM, RIKEN, and a grant from the autoimmune disease study group of Research in Intractable Diseases, Japanese Ministry of Health, Labor, and Welfare, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ykochi@src.riken.jp

☯ These authors contributed equally to this work.

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 $P<5.0\times10^{-8}$ (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 $P<1.0\times10^{-3}$ 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

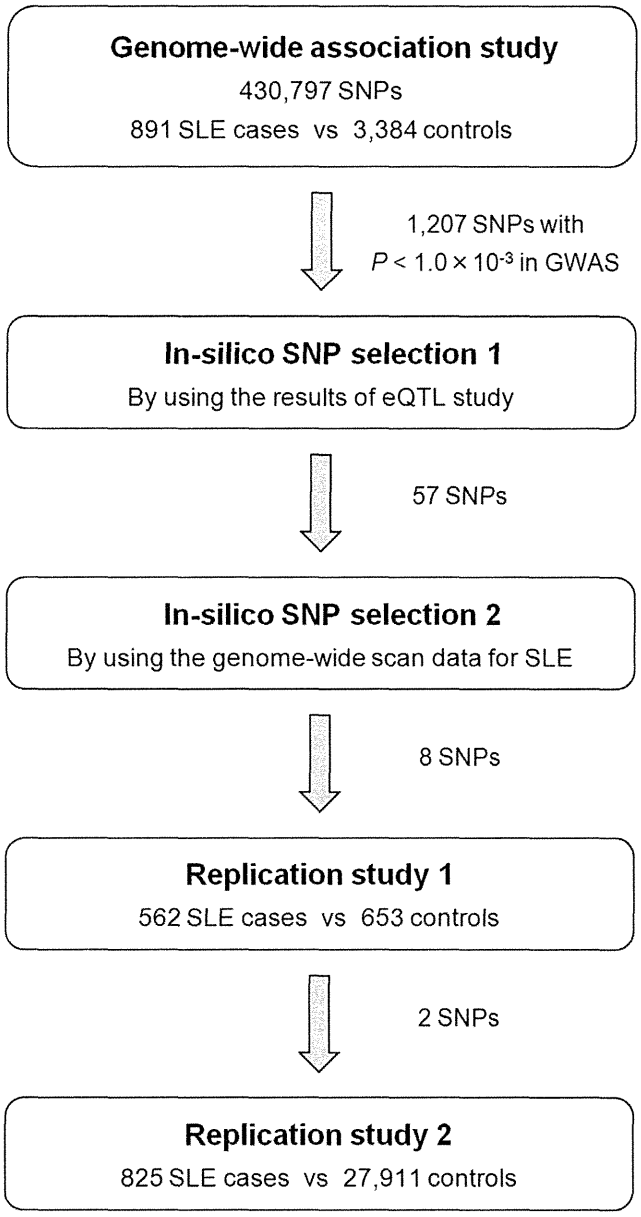

Figure 1. Design of the GWAS and multi-stage replication studies for SLE in Japanese subjects. A total of 2,278 SLE cases and 31,948 controls were enrolled. The clinical characteristics of the subjects are summarized in Table S1 and S2. Details of the genome-wide scan data for SLE referenced in the *in 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.

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

rsID[a]	Chr	Position (bp)	Cytoband	Gene	Allele[b]	No. subjects		Allele 1 freq.		OR (95%CI)	*P*
					1/2	Case	Control	Case	Control		
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 $P<5.0\times10^{-8}$ 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

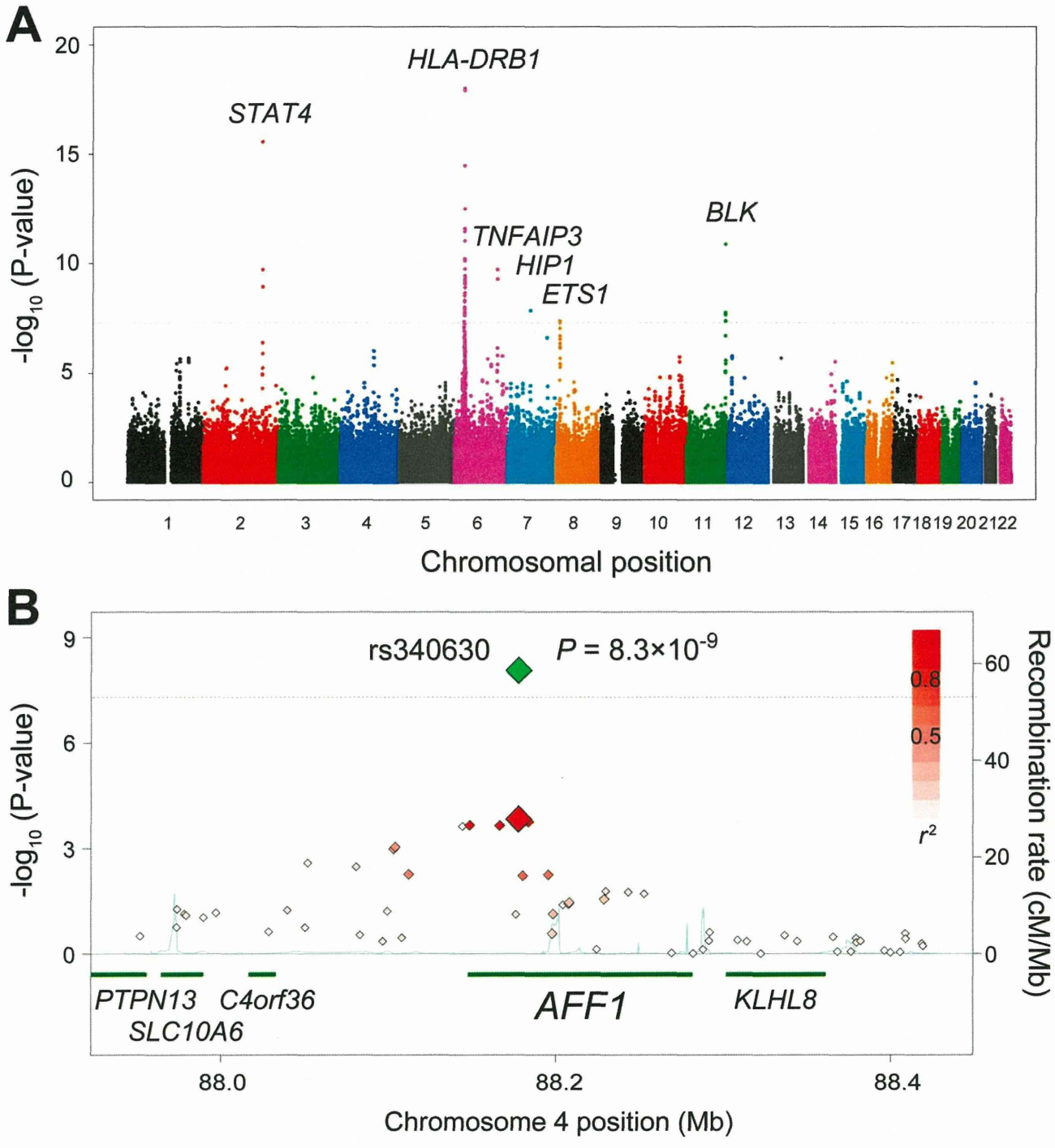

Figure 2. Associations of the *AFF1* locus with SLE. (A) A chromosomal plot of *P*-values in GWAS for SLE. (B) A regional plot in the *AFF1* locus. Diamond-shaped data points represent $-\log_{10}$ (*P*-values) of the SNPs. Large-sized points indicate the *P*-values of the landmark SNP, rs340630 (green 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 $P = 5.0 \times 10^{-8}$. 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.

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

rsID	Chr	Position (bp)	Cytoband	Gene	Allele[a]	Allele 1 freq.		OR (95%CI)	*P*	eQTL[b]	Identified by the studies in[c]	
					1/2	Case	Control				Caucasians	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 HumanHap550v3 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	Stage	No. subjects		Allele 1 freq.		OR (95%CI)	*P*	eQTL[a]
					1/2		Case	Control	Case	Control			
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

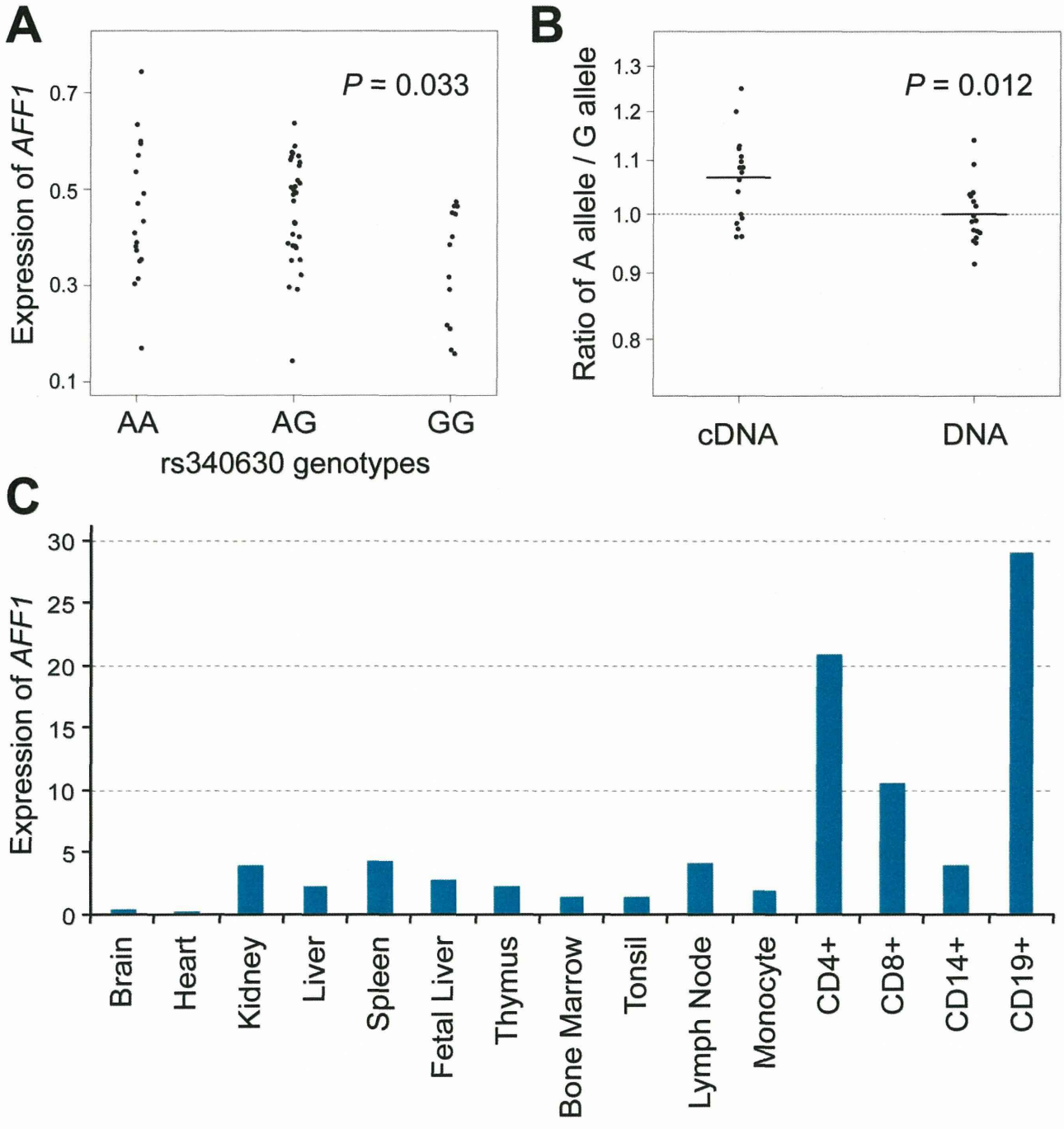

Figure 3. Association of rs340630 with *AFF1* expression. (A) Correlation between rs340630 genotypes and transcript levels of *AFF1* (NM_001166693) in EBV-transfected cell lines ($n = 62$) stimulated with PMA. (B) Allele-specific quantification (ASTQ) of *AFF1* transcripts. Allele specific-probes for 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 $<1.0\times10^{-6}$ 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′-CCCGGCGCAGTTTCTGAG-3′. The probe sequences were 5′-VIC-CGAAGACCGCCAGCGCCCAAC-TAMRA-3′ and 5′-FAM-CGAAGACCGCCGGCGCCCAA-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

eigenvectors 1 and 2 obtained from the PCA using EIGENSTRAT 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)

Acknowledgments

We are grateful to Dr. A. Miyatake, the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan, the autoimmune disease study group, and the staffs of the BioBank Japan Project for supporting the study and clinical sample collection. We thank the members of the Laboratory for Autoimmune Diseases, CGM, RIKEN, for their assistance.

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.

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.

Mod Rheumatol (2012) 22:52–58
DOI 10.1007/s10165-011-0467-y

ORIGINAL ARTICLE

A trans-ethnic genetic study of rheumatoid arthritis identified *FCGR2A* as a candidate common risk factor in Japanese and European populations

Roubila Meziani · Ryo Yamada · Meiko Takahashi · Kenei Ohigashi · Akio Morinobu · Chikashi Terao · Hitomi Hiratani · Koichiro Ohmura · Masao Yamaguchi · Takashi Nomura · Alexandre Vasilescu · Miki Kokubo · Victor Renault · Katsura Hirosawa · Chanavee Ratanajaraya · Simon Heath · Tsuneyo Mimori · Shimon Sakaguchi · Mark Lathrop · Inga Melchers · Shunichi Kumagai · Fumihiko Matsuda

Received: 28 February 2011 / Accepted: 22 April 2011 / Published online: 24 May 2011

Abstract Rheumatoid arthritis (RA) is a common systemic autoimmune disease and its onset and prognosis are controlled by genetic, immunological, and environmental factors. The *HLA* locus, particularly *HLA-DRB1*, is its strongest genetic risk determinant across ethnicities. Several other genes, including *PTPN22* and *PADI4*, show modest association with RA. However, they cover only a part of its genetic components and their relative contribution is different between populations. To identify novel genetic determinants, we took a candidate gene approach in a trans-ethnic manner. After critical selection of 169 genes based on their immunological function, we performed SNP discovery of these genes by the resequencing of exons and surrounding areas using European and Japanese DNAs. We then generated a panel of 1,509 SNPs for case–control association study in both populations. The DerSimonian–Laird test for meta-analysis, using the combined results of the two populations, identified rs7551957 at the 5′-flanking region of the low-affinity Fc-gamma receptor IIa (*FCGR2A*) gene as the strongest candidate for the association ($p = 8.6 \times 10^{-5}$, odds ratio = 1.58 with 95%CI 1.25–1.99). Suggestive signals were also obtained for three SNPs in the dihydropyrimidine dehydrogenase (*DPYD*) gene (rs6685859; $p = 1.3 \times 10^{-4}$, rs7550959; $p = 1.5 \times 10^{-4}$ and rs7531138; $p = 1.7 \times 10^{-4}$) and an intronic SNP, rs2269310, of the erythrocytic spectrin beta (*SPTB*) gene ($p = 7.9 \times 10^{-4}$).

Electronic supplementary material The online version of this article (doi:10.1007/s10165-011-0467-y) contains supplementary material, which is available to authorized users.

R. Meziani · H. Hiratani · M. Yamaguchi · M. Kokubo · V. Renault · K. Hirosawa · F. Matsuda
CREST Program, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan

R. Meziani · S. Heath · M. Lathrop · F. Matsuda
Centre National de Genotypage, Institute Genomique, Commissariat a l'Energie Atomique, 91057 Evry, France

R. Meziani · M. Lathrop
Fondation Jean Dausset-CEPH, 75010 Paris, France

R. Yamada · M. Takahashi · K. Ohigashi · C. Terao · H. Hiratani · M. Yamaguchi · A. Vasilescu · M. Kokubo · V. Renault · K. Hirosawa · C. Ratanajaraya · F. Matsuda (✉)
Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Kyoto 606-8501, Japan
e-mail: fumi@genome.med.kyoto-u.ac.jp

A. Morinobu · S. Kumagai
Department of Clinical Pathology and Immunology, Graduate School of Medicine, Kobe University, Kobe 650-0017, Japan

C. Terao · K. Ohmura · T. Mimori
Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan

T. Nomura · S. Sakaguchi
Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan

A. Vasilescu · F. Matsuda
Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University, Kyoto 606-8501, Japan

I. Melchers
Clinical Research Unit for Rheumatology, Department of Rheumatology and Clinical Immunology, University Medical Center Freiburg, 79106 Freiburg, Germany

Keywords *FCGR2A* · Genotyping · Rheumatoid arthritis · Single nucleotide polymorphism · Trans-ethnic study

Introduction

Rheumatoid arthritis (RA [OMIM: 180300]) is a systemic autoimmune disease [1], and is one of the most common forms of inflammatory arthritis, affecting up to 1% of the adult population [2]. Symptoms are chronic, destructive, and debilitating arthritis with a variation in the number of clinical features, such as the presence of autoantibody and joint erosions [3]. Clinical manifestation of RA is related to the development of a variety of autoantibodies, including antibodies to citrullinated peptide antigens and rheumatoid factor, although their pathological role is still unclear [4, 5].

Compelling evidence from genome-wide association (GWA) studies demonstrated that the *HLA* locus is the strongest genetic determinant beyond ethnicity [6]. However, the *HLA* locus contributes to only approximately 37% of the overall genetic susceptibility [7], suggesting the presence of other genes that are genetically associated with pathogenesis, clinical phenotype, and disease heterogeneity. Moreover, the relative contribution of RA-related genes is considered to be different between ethnicities. Indeed, the *PTPN22* gene was shown to be associated with RA in populations of European descent, but not in Asians [8]. Similarly, the *PADI4* gene showed a strong association with RA in Asians [9], but the association was much weaker in Europeans. Intronic SNPs in the *STAT4* gene were identified to be strongly associated with RA in Europeans [10], and their modest statistical association was confirmed in Asians [11], providing the first example of non-*HLA* RA-associated genes in two major ethnicities.

Importantly, however, very few hypothesis-independent GWA studies have succeeded in the identification of non-*HLA* genes associated with RA. A genome scan by the Wellcome Trust study was only able to successfully identify *HLA* and *PTPN22* loci as genome-wide significant [6]. This may be due to the disease heterogeneity with phenocopies, and the insufficient power to detect genes with modest effects [12]. Indeed, SNPs associated with RA in non-*HLA* genes showed only intermediate allelic odds ratios (ORs) of 1.3–2.0. To overcome such shortcomings, studies need to be enlarged, enrolling several thousand patients and a similar number of controls. Alternatively, a candidate gene approach, which focuses on genes critically selected by their biological function, can be an efficient and more cost-effective strategy because SNPs that are associated with RA are mostly within genes that are functionally implicated in the pathogenicity of RA [13]. Above all, comparative genetic analysis across different ethnic groups is ideal.

In this study, we performed a trans-ethnic case–control association study of RA in Japanese and European RA patients by applying an SNP marker panel of 169 genes related to the immune system and drug metabolisms designed for genetic approaches of various immune-related diseases. A total of 1,509 SNPs were exhaustively genotyped in 238 RA cases and 184 controls of Japanese origin, as well as in 182 cases and 273 controls of German and French populations.

Materials and methods

Study populations

Sample collection and genotyping was initiated after approval from the local ethical committees. Written informed consent was obtained from all patients and control individuals after adequate explanation of the study. The disease subjects were recruited in Germany and in Japan. The 420 patients with RA consisted of 182 Germans from the Freiburg area and 238 Japanese from the Kyoto and Kobe areas. All subjects satisfied the American Rheumatism Association's revised criteria (1987) for classification of RA [14]. For population controls, 91 German subjects from the Freiburg area and 184 Japanese subjects were recruited. Additionally, 182 French Caucasian controls from the Epidemiological study on the Genetics and Environment of Asthma (EGEA) were used.

Selection of candidate genes and single nucleotide polymorphisms

A total of 169 genes on autosomes were chosen for case–control association studies of different immune-related diseases, such as autoimmune diseases, immune deficiencies, and allergies, as well as drug response in such diseases. The selection was made according to their biological function in the immune system, such as cytokines and their receptors, cell adhesion molecules, transcriptional factors, and genes involved in signal transduction and cell–cell communication (Table 1). Genes already known for their association with the diseases, in addition to those shown to be related to drug metabolism prior to 2006 when the SNP discovery was performed, were also included in the panel.

The identification of SNPs was performed by an exhaustive resequencing of exons and flanking regions of these genes, using 32 each of French and Japanese control DNAs. Additional SNPs located in the linkage disequilibrium (LD) blocks covering the 169 genes in the International HapMap Project were also included. Among approximately 10,000 SNPs, those having a minor allele frequency (MAF) greater than 0.05 were tested using the

Table 1 Classification of candidate genes by their immunological function

Signal transduction	*ABCB1, ANK1, APOH, CARD4, CARD15, CBLB, CCND2, CCND2P, CCND3, CCNDBP1, CDC37, CDKN1A, CDKN2A2, EIF2AK2, EVI5L, GAB1, GAB2, GRB2, HCK, HSPCA, HSPCB, ITPKB, KIR3DL2, LAT, LCK, LCP2, MBL2, PTPN11, PTPN13, RASGRP1, SOCS3, SOS1, SYK, TANK, TGFB1, VAV1, VAV2, VAV3, XPO1, ZAP70, FTH1, NPM1*
Transcription factors	*AIRE, APOBEC3G, CCNT1, CCNT2, CDK7, CDK9, CD3EAP, NFAT5, NFATC1, NFATC2, NFATC3, NFATC4, STAT1, TBX1, TGFB1I1, TRIM21, TROVE2, TSC22D1, YBX1*
Cell–cell communication	*CD274, CD28, CD36, CD38, CD3Z, CD4, CD58, CD8A, CD8B, COMMD1, CTLA4, FCGR2A, FCGR3A, FER, ICAM2, ICAM3, ITGA1, ITGA2, ITGA4, ITGAL, ITGAV, ITGB1, ITGB4, PECAM1, SELL, TGFB, VCAM1, VIL2, VILL*
Cytokines and receptors	*FAS, FASLG, IFNA2, IFNAR1, IFNG, IFNGR1, IFNGR2, IL10, IL10RA, IL10RB, IL12A, IL12B, IL13, IL15, IL18, IL1A, IL1B, IL1R1, IL1RN, IL2, IL21R, IL2RA, IL2RB, IL4, IL4R, IL6, IL6R, IL6ST, IL7, IL8, L8RA, IL8RB, KIT, TNFRSF11A, TNFRSF13C, TNFRSF18, TNFRSF1A, TNFRSF1B, TNFRSF7, TNFSF15, TNFSF7*
Metabolism	*ALDH2, ANPEP, BCHE, CYP2C19, CYP2D6, CYP2E1, DPP4, DPYD, EPB41, FKBP4, GSTM4, GSTP1, NAT2, NOS2A, NQO1, NQO2, PLCG1, PPIA, SOD1, SOD2, SPTA1, SPTB, TPMT, TRIM5*
Regulation factor of immune response	*GYPA, GYPB, GYPC, OAS1, OAS2, OAS3, THY1*
Genes in the *HLA* locus	*HLA-DOB, PSMB8, PSMB9, TAP2, LTA, CSNK2B*
Other	*C1QBP*: role in complement pathway (binding C1q protein)

GoldenGate technology marker panel selection program by Illumina Inc. (San Diego, CA, USA), and only markers with Illumina Design Scores (IDS) greater than 0.4 were chosen. SNPs in the same LD block with pairwise-r^2 greater than 0.8 were divided into subgroups, and the SNP which showed the highest IDS in each subgroup was selected. When there were multiple SNPs sharing the highest IDS, the one with the highest MAF was chosen. Following these selection steps, a total of 1,509 SNPs were finally chosen as tag SNP markers for the genotyping study.

Genotyping and quality control

Genomic DNAs were extracted from fresh peripheral blood mononuclear cells or from EBV-transformed lymphoblasts in accordance with protocols approved by the appropriate authority.

A panel of 1,509 SNPs was genotyped using a GoldenGate assay on an Illumina BeadArray genotyping platform according to the manufacturer's instructions. DNA samples were tracked using a laboratory information management system (LIMS), and genotypes were called using the Genotyping module of BeadStudio 2 software (Illumina Inc.). The results obtained were filtered on the basis of genotype call rates (success rates of >90% for marker, >95% for DNA sample).

Statistical analysis

Genotype distribution was evaluated by the Hardy–Weinberg equilibrium in the control group (χ^2 test), and the markers with p values less than 0.05 were excluded from the tests [15]. Allele frequency of each SNP was compared for association with RA between cases and controls in each population using the trend χ^2 test and a non-biased exact trend test [16], as well as the DerSimonian–Laird test for meta-analysis, using the combined results of two populations [17].

Expression analyses of the FCGR gene family

A gene-expression dataset in lymphoblastoid cell lines derived from 210 unrelated HapMap populations (GSE6536) was obtained from the Gene Expression Omnibus (GEO) database [18]. The correlation between the expression data of *FCGR1A*, *2A*, *2B*, *2C*, *3A*, and *3B*, and rs7551957 genotypes of 268 individuals (89 European, 44 Japanese, 45 Chinese, and 90 West African Yoruba) available from the HapMap phase 2 data, was examined using the calculation program recommended by the GEO. The association p values were obtained by the Joncheere–Terepstra method using R software or SPSS (version 18).

Results

Association analysis in Japanese and European populations

In order to identify genetic loci associated with susceptibility to RA, 1,509 SNP markers representing 169 candidate genes with an average number of 8.9 SNPs (ranging from 1 to 75) per gene, were genotyped in DNA samples of Japanese and European origins (summarized in Table 1). After quality control of the results, 1,375 and 1,330 SNPs in the Japanese and Europeans, respectively, were tested for association.

In the Japanese case–control analysis, a total of 41 SNPs tagging 26 genes showed a nominal p value <0.005, of which 13 SNPs belonging to ten genes were less than 0.001 (Supplementary Table 1). Of those 41 SNPs, rs17587 in exon4 of the *PSMB9* gene was non-synonymous (arginine to histidine), while the others were either in the gene-flanking regions (11 SNPs), 3′-untranslated region (one SNP) or in introns (28 SNPs). The association analysis using the European sample set identified a putative difference in allele frequency of 20 SNPs in 16 genes with a nominal p value <0.005, of which five SNPs corresponding to three genes showed a p value <0.001 (Supplementary Table 2). One synonymous SNP, rs2302872, was located in exon14 of the *DPP4* gene, while the others were either in the gene-flanking regions (11 SNPs) or in introns (nine SNPs). Markers that are located in the *HLA* locus showed association with RA in both populations. The strongest p value was observed for rs1894408, located adjacent to the *HLA-DOB* gene ($p = 7.5 \times 10^{-5}$) in the Japanese, and for rs1383266 near the *PSMB9* gene ($p = 0.0011$) in the Europeans (Supplementary Tables 1 and 2).

Meta-analysis using combined results of Japanese and European studies

To identify the genes/variants that are associated with RA in both Japanese and Europeans, we performed a meta-analysis combining the genotyping results of the two populations by using the DerSimonian–Laird test. In total, 22 SNPs corresponding to 15 genes showed a p value (*DLp*) less than 0.01 (Table 2). Among them, five SNPs, namely, rs6685859, rs7551957, rs2269304, rs4819522, and rs5746834, showed relatively modest association with RA in both populations (p = <0.05).

The strongest association was obtained for rs7551957 of the *FCGR2A* (low affinity Fc-gamma receptor IIa) gene, with $DLp = 8.6 \times 10^{-5}$ and OR = 1.58 with 95% confidence interval (CI) 1.25–1.99. This SNP showed a strong association with RA in both populations (trend χ^2 p = 0.0035 in the Japanese and $p = 0.0062$ in the Europeans). Three suggestive signals were obtained in the *DPYD* (dihydropyrimidine dehydrogenase) gene by meta-analysis: rs6685859 ($DLp = 1.3 \times 10^{-4}$ and OR = 1.64 with CI 1.27–2.12), rs7550959 ($DLp = 1.5 \times 10^{-4}$ and OR = 1.52 with CI 1.52–1.89), rs7531138 ($DLp = 1.7 \times 10^{-4}$ and OR = 1.51 with CI 1.22–1.88; Table 2). One SNP, rs2269310, in intron1 of the *SPTB* (erythrocytic spectrin beta) gene also showed a modest association ($DLp = 7.9 \times 10^{-4}$ and OR = 1.51 with CI 1.18–1.92) with RA.

Discussion

To our knowledge, this is the first report of a candidate gene-based trans-ethnic association analysis of RA. A panel of 169 genes that play important roles in immune response was extensively studied by SNP genotyping in Japanese and European populations. The *HLA* gene cluster is the major gene locus that contributes to RA susceptibility, and the *HLA-DRB1* gene is reported to be the strongest candidate [6]. In our study, the association between SNPs in the *HLA* region and RA was confirmed in both populations when the results were analyzed separately. However, these SNPs were not found to be significant in the meta-analysis. It is well known that the LD structure in this region is markedly different between ethnic groups. Hence, a larger sample size is required for sufficient statistical power to detect the same SNP as significant in both populations. Moreover, the panel used in this study only contains a single *HLA* gene, *HLA-DOB*, and not the others that are proven to be associated with RA.

The strongest association was obtained in the *FCGR2A* gene, which is located within the cluster of low affinity Fc-gamma receptor genes (*FCGR2A*, *FCGR3A*, *FCGR2C*, *FCGR3B*, and *FCGR2B*) on chromosome 1q22-23 [19]. Rs7551957 showed a strong association with RA in the two ethnicities ($p = 0.0035$ in the Japanese and $p = 0.0062$ in the Europeans), which was further confirmed by meta-analysis ($DLp = 8.6 \times 10^{-5}$;Table 2). In both populations, the variant allele frequency is higher in the RA cases (0.170 in the Japanese and 0.466 in the Europeans) compared with the controls (0.098 in the Japanese and 0.372 in the Europeans). Another SNP in the *FCGR2A* gene, rs1801274, showed a modest association in the meta-analysis ($DLp = 0.0021$). The difference in trend of p value between the two populations ($p = 0.098$ in the Japanese and $p = 0.011$ in the Europeans) may well be explained by the difference in the allele frequencies (0.225 in the Japanese and 0.500 in the Europeans). In line with our results, a recent meta-analysis of European genome-wide association studies confirmed the association of *FCGR2A* with RA risk ($p = 0.0004$) [20]. FCGRs are expressed on the surface of cells involved in the immune

Table 2 Single nucleotide polymorphisms associated with rheumatoid arthritis in meta-analysis using the combined results of Japanese and European populations

dbSNP ID	Chr.	Gene	Location	Nucleotide		Amino acid		Japanese freq. A2		Nominal trend *p*	European freq. A2		Nominal trend *p*	Meta-analysis	
				Ref. (A1)	Var. (A2)	Ref.	Var.	Case	Cont		Case	Cont		*DLp*	OR (95%CI)
rs6685859	1p21.3	*DPYD*	intron16	G	C			0.023	0.050	0.042	0.506	0.619	5.1×10^{-4}	1.3×10^{-4}	1.64 (1.27–2.12)
rs7550959			intron13	G	A			0.166	0.215	0.079	0.387	0.504	6.4×10^{-4}	1.5×10^{-4}	1.52 (1.22–1.89)
rs7531138			intron13	T	A			0.168	0.215	0.090	0.392	0.509	5.9×10^{-4}	1.7×10^{-4}	1.51 (1.22–1.88)
rs7551957	1q23.3	*FCGR2A*	5′-flanking	T	C			0.098	0.170	0.0035	0.372	0.466	0.0062	8.6×10^{-5}	1.58 (1.25–1.99)
rs1801274			exon4	A	G	His	Arg	0.176	0.225	0.098	0.412	0.500	0.011	0.0021	1.40 (1.13–1.74)
rs697846	1q42.12	*ITPKB*	3′-flanking	A	G			0.202	0.283	0.010	0.140	0.181	0.14	0.0042	1.44 (1.12–1.85)
rs1050567	2p15	*XPO1*	3′UTR	C	T			0.236	0.331	0.0053	0.102	0.133	0.17	0.0021	1.50 (1.15–1.94)
rs3770768	2p22.2	*EIF2AK2*	intron13	G	A			0.004	0.014	0.15	0.099	0.155	0.015	0.0082	1.72 (1.15–2.58)
rs926169	2q33.2	*CTLA4*	5′-flanking	G*	T			0.660	0.597	0.091	0.423	0.357	0.047	0.0087	1.31 (1.07–1.61)
rs2686399	3q.26.1	*BCHE*	3′-flanking	C*	G			0.844	0.819	0.36	0.732	0.648	0.0064	0.0077	1.37 (1.08–1.73)
rs6946119	7q21.12	*ABCB1*	3′-flanking	T	C			0.121	0.159	0.12	0.238	0.304	0.028	0.0076	1.39 (1.09–1.78)
rs2269310	14q23.3	*SPTB*	intron26	G	A			0.368	0.475	0.0054	0.108	0.147	0.073	7.9×10^{-4}	1.51 (1.18–1.92)
rs229670			intron1	A*	C			0.413	0.349	0.061	0.268	0.177	0.0011	0.0024	1.48 (1.14–1.91)
rs2269304			intron14	C*	A			0.118	0.075	0.049	0.213	0.153	0.018	0.0025	1.54 (1.16–2.05)
rs4787426	16p12.1	*IL4R*	3′-flanking	T	G			0.104	0.170	0.012	0.122	0.153	0.21	0.0076	1.49 (1.11–2.01)
rs4968681	17q23.3	*ICAM2*	5′-flanking	G	A			0.323	0.409	0.016	0.326	0.371	0.17	0.0082	1.32 (1.07–1.62)
rs7503550	17q23.3	*PECAM1*	intron4	G*	A			0.529	0.467	0.096	0.566	0.474	0.0072	0.0018	1.37 (1.12–1.67)
rs537188	19p13.2	*EVI5L*	intron11	G	A			0.085	0.105	0.34	0.140	0.212	0.0032	0.0045	1.52 (1.13–2.04)
rs347033	19p13.3	*VAV1*	intron4	T	C			0.144	0.227	0.0045	0.163	0.205	0.11	0.0028	1.50 (1.15–1.97)
rs2866370	20q12	*PLCG1*	intron1	G	A			0.039	0.077	0.022	0.050	0.078	0.080	0.0054	1.81 (1.19–2.75)
rs4819522	22q11.21	*TBX1*	exon9	C	T	Thr	Met	0.058	0.097	0.042	0.182	0.254	0.013	0.0014	1.58 (1.19–2.09)
rs5746834			3′-flanking	G	T			0.037	0.069	0.039	0.190	0.248	0.039	0.0049	1.52 (1.13–2.03)

SNPs with a *p* value (*DLp*) less than 0.01 according to the DerSimonian–Laird test are listed with odds ratio (OR) and 95% confidence interval (95%CI).

The risk allele is indicated by an asterisk if it is the reference allele.

Arg arginine, *His* histidine, *Met* methionine, *Thr* threonine, *Ref.* reference allele, *Var.* variant allele

system, and participate in diverse functions such as phagocytosis of immune complexes and modulation of antibody production by B cells. Various genetic polymorphisms of these receptors were reported to be associated with several autoimmune diseases [21, 22]. In particular, *FCGR2A* was shown to be associated to systemic lupus erythematosus [23, 24]. In the mouse model, *Fcgr3*-deficient hosts exhibit resistance to arthritis induced by collagen type II or anti-glucose-6-phosphate isomerase antibody [25]. In contrast, mice deficient for *Fcgr2b* lead to increased susceptibility to collagen-induced arthritis [26]. Theses findings suggest that expression of FCGRs on synovial cells may contribute to the antibody-triggered inflammation in joints [27]. In addition, we examined the association of rs7551957 with the expression level of the other *FCGR* family members according to the four population groups (European, Japanese, Chinese, and West African Yoruba), but linear regression analyses failed to reveal any significant associations (results not shown).

Variation in gene copy number is postulated to influence clinical phenotype. There have been conflicting reports regarding the association of copy number variations (CNV) in the *FCGR* locus with RA in Caucasian studies [28, 29]. In the vicinity of the *FCGR* locus there are at least three reported regions showing CNV, but rs7551957 is located more than 19-kb away from the nearest CNV region, which extends from the 3′-UTR of *FCGR2A* to the 3′-UTR of *FCGR2C* [30]. In addition, a careful examination of the rs7551957 genotype results did not reveal any indications of CNV in the observed cluster signals. Hence the association observed for rs7551957 with *FCGR2A* is unlikely to be caused by CNV of the *FCGR* locus.

Rs6685859 in intron16 of the *DPYD* gene showed $DLp = 1.3 \times 10^{-4}$, although trend p value was at a marginal level ($p = 0.042$) in the Japanese compared with the Europeans ($p = 5.1 \times 10^{-4}$). Again this may be due to the difference in allele frequencies (case versus control; 0.023 versus 0.050 in the Japanese and 0.506 versus 0.619 in the Europeans), because the risk allele is the same in both populations. The two other markers, rs7550959 and rs7531138, showed similar trends as rs6685859. DPYD is a pyrimidine catabolic enzyme, mutation of which leads to dihydropyrimidine dehydrogenase deficiency, putting cancer patients receiving 5-fluorouracil chemotherapy at an increased risk of toxicity [31]. To our knowledge, there have been no published reports on its association with RA or other autoimmune-related diseases to date.

Among the SNPs in the *SPTB* gene examined, three (rs2269310, rs229670, and rs2269304) showed significant *DLp* values. Rs2269310, which showed $DLp = 7.9 \times 10^{-4}$ and a significant trend p value of 0.0054 in the Japanese, did not show a similar level of association in the Europeans ($p = 0.073$). Again this may be due to a lower frequency of the variant allele in Europeans. SPTB acts to stabilize erythrocyte membranes, and mutations in the *SPTB* gene have been associated with spherocytosis type 2 [32], and neonatal hemolytic anemia [33]. Again, its association with RA is not known, although it has been previously reported that there was no significant immunoreactivity observed against spectrin in the sera of 50 RA patients [34].

In the current study, none of the SNPs remained significant after nominal trend p values were corrected for multiple testing using the Bonferroni method (data not shown). Therefore, detection of significant association requires replication analyses using independent sample sets. Nonetheless, we succeeded in the identification of several genetic variants as being associated with RA susceptibility across ethnicities, indicating the usefulness of a trans-ethnic comparison. Additional research is necessary to further confirm their association with the disease, and to elucidate their biological role in the pathophysiology of RA.

Acknowledgments The authors are grateful to all the patients and medical staff who have kindly collaborated on this project and also to the Epidemiological Study on the Genetics and Environment of Asthma (EGEA) cooperative group, who allowed us access to data on the EGEA study. This work was supported in part by a grant from the Federal Ministry of Education and Research (Competence Network Systemic Inflammatory Rheumatic Diseases) to I. M., and by the CREST program from the Japan Science and technology Agency (Saitama, Japan).

Conflict of interest None.

References

1. Firestein GS. Evolving concepts of rheumatoid arthritis. Nature. 2003;423:356–61.
2. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res. 2002;4(Suppl 3):S265–72.
3. Choy EH, Panayi GS. Cytokine pathways and joint inflammation in rheumatoid arthritis. N Engl J Med. 2001;344:907–16.
4. van Zeben D, Hazes JM, Zwinderman AH, Cats A, van der Voort EA, Breedveld FC. Clinical significance of rheumatoid factors in early rheumatoid arthritis: results of a follow-up study. Ann Rheum Dis. 1992;51:1029–35.
5. Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest. 1998;101:273–81.
6. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–78.
7. Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. Clin Genet. 1989;36:178–82.
8. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase

(PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet. 2004;75:330–7.
9. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet. 2003;34:395–402.
10. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med. 2007;357:977–86.
11. Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. Mol Med. 2007;13:455–60.
12. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6: 95–108.
13. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet. 2005;77:1044–60.
14. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31:315–24.
15. Gomes I, Collins A, Lonjou C, Thomas NS, Wilkinson J, Watson M, et al. Hardy–Weinberg quality control. Ann Hum Genet. 1999;63:535–8.
16. Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered. 2002;53:146–52.
17. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88.
18. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315: 848–53.
19. Oakey RJ, Howard TA, Hogarth PM, Tani K, Seldin MF. Chromosomal mapping of the high affinity Fc gamma receptor gene. Immunogenetics. 1992;35:279–82.
20. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010;42:508–14.
21. Dijstelbloem HM, Scheepers RH, Oost WW, Stegeman CA, van der Pol WL, Sluiter WJ, et al. Fcgamma receptor polymorphisms in Wegener's granulomatosis: risk factors for disease relapse. Arthritis Rheum. 1999;42:1823–7.
22. Myhr KM, Raknes G, Nyland H, Vedeler C. Immunoglobulin G Fc-receptor (FcgammaR) IIA and IIIB polymorphisms related to disability in MS. Neurology. 1999;52:1771–6.
23. Moser KL, Neas BR, Salmon JE, Yu H, Gray-McGuire C, Asundi N, et al. Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African–American pedigrees. Proc Natl Acad Sci USA. 1998;95:14869–74.
24. Manger K, Repp R, Jansen M, Geisselbrecht M, Wassmuth R, Westerdaal NA, et al. Fc gamma receptor IIa, IIIa, and IIIb polymorphisms in German patients with systemic lupus erythematosus: association with clinical symptoms. Ann Rheum Dis. 2002;61:786–92.
25. Diaz de Stahl T, Andren M, Martinsson P, Verbeek JS, Kleinau S. Expression of FcgammaRIII is required for development of collagen-induced arthritis. Eur J Immunol. 2002;32:2915–22.
26. Kleinau S, Martinsson P, Heyman B. Induction and suppression of collagen-induced arthritis is dependent on distinct fc gamma receptors. J Exp Med. 2000;191:1611–6.
27. Magnusson SE, Engstrom M, Jacob U, Ulfgren AK, Kleinau S. High synovial expression of the inhibitory FcgammaRIIb in rheumatoid arthritis. Arthritis Res Ther. 2007;9:R51.
28. Marques RB, Thabet MM, White SJ, Houwing-Duistermaat JJ, Bakker AM, Hendriks GJ, et al. Genetic variation of the Fc gamma receptor 3B gene and association with rheumatoid arthritis. PLoS One. 2010;5.
29. McKinney C, Fanciulli M, Merriman ME, Phipps-Green A, Alizadeh BZ, Koeleman BP, et al. Association of variation in Fc gamma receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples. Ann Rheum Dis. 2010;69:1711–6.
30. Niederer HA, Willcocks LC, Rayner TF, Yang W, Lau YL, Williams TN, et al. Copy number, linkage disequilibrium and disease association in the FCGR locus. Hum Mol Genet. 2010; 19:3282–94.
31. Van Kuilenburg AB, van Lenthe H, Blom MJ, Mul EP, Van Gennip AH. Profound variation in dihydropyrimidine dehydrogenase activity in human blood cells: major implications for the detection of partly deficient patients. Br J Cancer. 1999;79:620–6.
32. Goodman SR, Shiffer KA, Casoria LA, Eyster ME. Identification of the molecular defect in the erythrocyte membrane skeleton of some kindreds with hereditary spherocytosis. Blood. 1982;60: 772–84.
33. Eber SW, Morris SA, Schroter W, Gratzer WB. Interactions of spectrin in hereditary elliptocytes containing truncated spectrin beta-chains. J Clin Invest. 1988;81:523–30.
34. Shrivastav M, Mittal B, Aggarwal A, Misra R. Autoantibodies against cytoskeletal proteins in rheumatoid arthritis. Clin Rheumatol. 2002;21:505–10.