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.

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

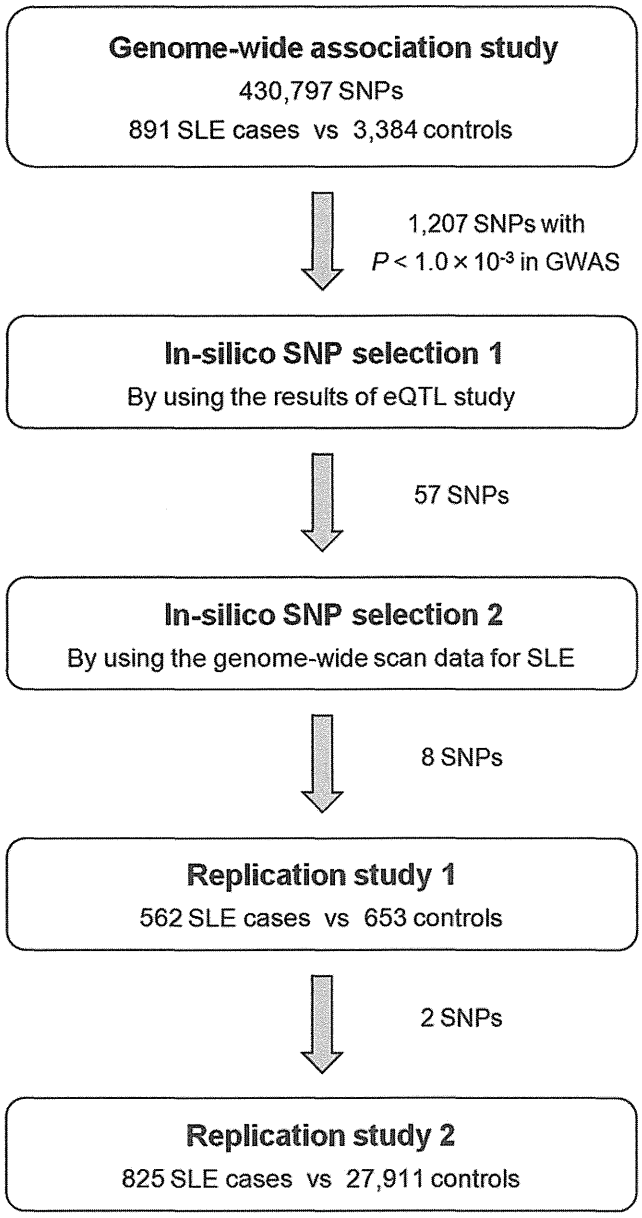

Figure 1. Design of the GWAS and multi-stage replication studies for SLE in Japanese subjects. A total of 2,278 SLE cases and 31,948 controls were enrolled. The clinical characteristics of the subjects are summarized in Table S1 and S2. Details of the genome-wide scan data for SLE referenced in the *in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

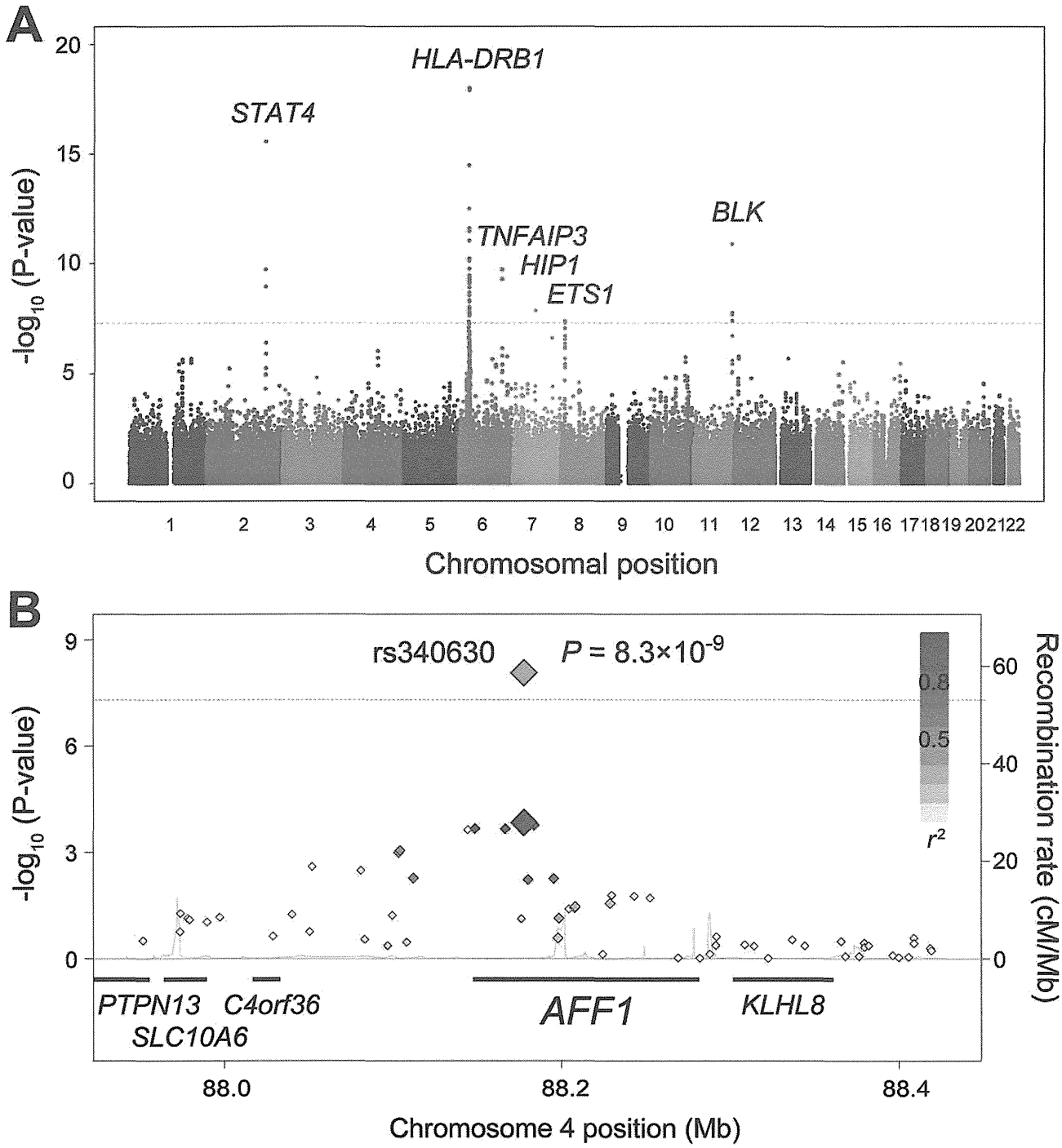

Figure 2. Associations of the *AFF1* locus with SLE. (A) A chromosomal plot of *P*-values in GWAS for SLE. (B) A regional plot in the *AFF1* locus. Diamond-shaped data points represent $-\log_{10}$ (*P*-values) of the SNPs. Large-sized points indicate the *P*-values of the landmark SNP, rs340630 (green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

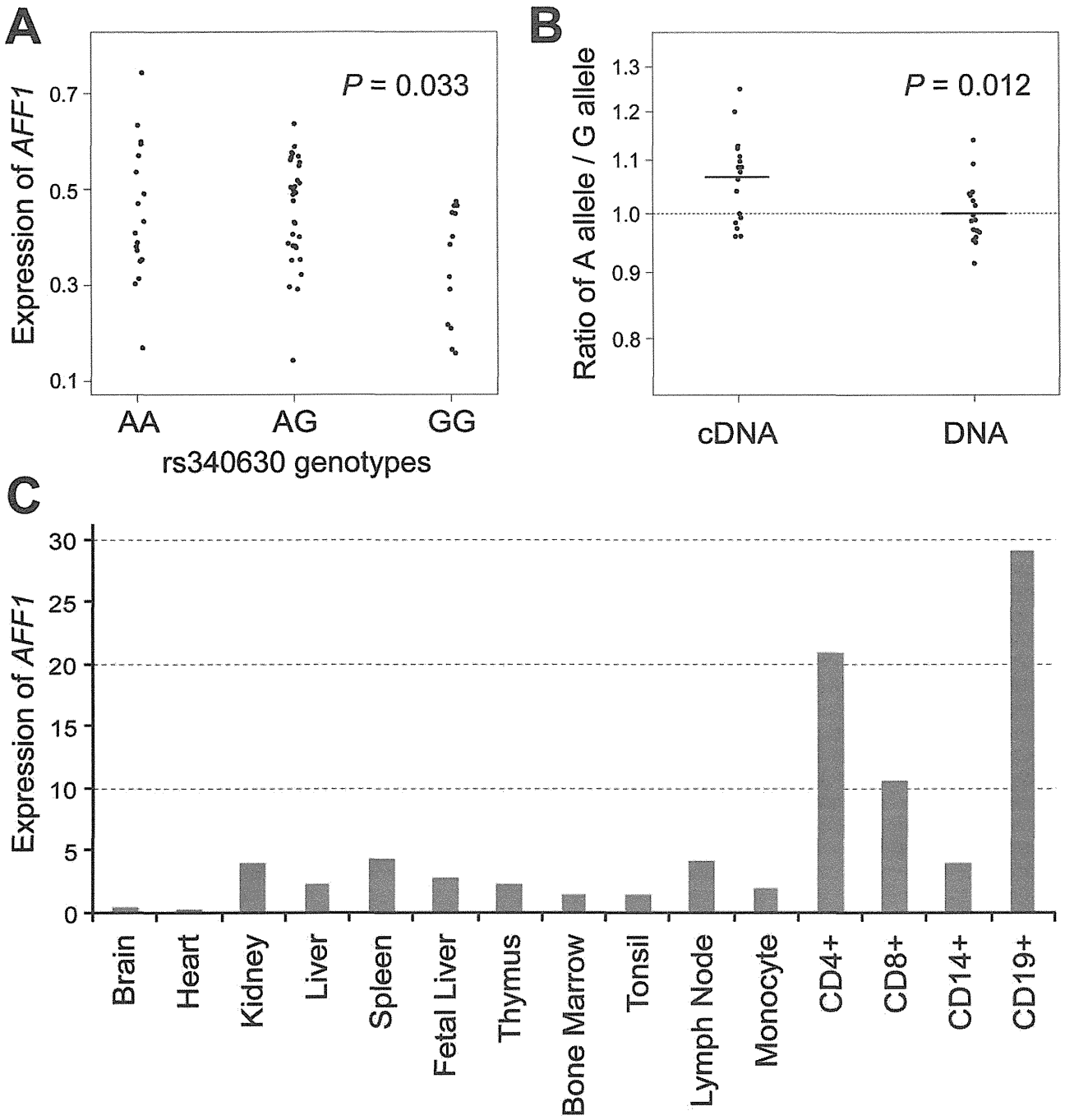

Figure 3. Association of rs340630 with *AFF1* expression. (A) Correlation between rs340630 genotypes and transcript levels of *AFF1* (NM_001166693) in EBV-transfected cell lines ($n = 62$) stimulated with PMA. (B) Allele-specific quantification (ASTQ) of *AFF1* transcripts. Allele specific-probes for 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 *AFF1* 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′-CCCGGCGCA-GTTTCTGAG-3′. The probe sequences were 5′-VIC-CGAA-GACCGCCAGCGCCCAAC-TAMRA-3′ and 5′-FAM-CGAA-GACCGCCGGCGCCCAA-TAMRA-3′. Ct values of VIC and FAM were obtained for genomic DNA and cDNA samples after 40 cycles of real-time PCR. We also prepared genomic DNA of samples homozygous for each allele and mixed them at different ratios (2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2) to create a standard curve by plotting Ct values of VIC/FAM against the allelic ratio of VIC/FAM for each mixture. Using the standard curve, we calculated the allelic ratios for each genomic DNA and cDNA samples. We measured each sample in quadruplicate in one assay; tests were independently repeated twice.

Web resources

The URLs for data presented herein are as follows.

NCBI GEO, http://www.ncbi.nlm.nih.gov/geo

BioBank Japan Project, http://biobankjp.org

PLINK software, http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml

International HapMap Project, http://www.hapmap.org

EIGENSTRAT software, http://genepath.med.harvard.edu/~reich/Software.htm

MACH and mach2qtl software, http://www.sph.umich.edu/csg/abecasis/MACH/index.html

SNAP, http://www.broadinstitute.org/mpg/snap/index.php

Supporting Information

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

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.

Journal of Medical Virology 84:61–70 (2012)

Predictive Value of Early Viral Dynamics During Peginterferon and Ribavirin Combination Therapy Based on Genetic Polymorphisms Near the *IL28B* Gene in Patients Infected With HCV Genotype 1b

Hidenori Toyoda,[1]* Takashi Kumada,[1] Toshifumi Tada,[1] Kazuhiko Hayashi,[2] Takashi Honda,[2] Yoshiaki Katano,[2] Hidemi Goto,[2] Takahisa Kawaguchi,[3] Yoshiki Murakami,[3] and Fumihiko Matsuda[3]

[1]*Department of Gastroenterology, Ogaki Municipal Hospital, Ogaki, Japan*
[2]*Department of Gastroenterology, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[3]*Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan*

A study was carried out to determine whether early viral dynamics retain prediction of the outcome of peginterferon (PEG-IFN) and ribavirin combination therapy based on different genetic polymorphisms near the *IL28B* gene, the strongest baseline predictor of response to this therapy. A total of 272 patients infected with hepatitis C virus (HCV) genotype 1b were grouped according to genetic polymorphisms near the *IL28B* gene (rs8099917). The ability of reduced HCV RNA levels at 4 and 12 weeks after starting therapy to predict a sustained virologic response was evaluated based on these genotypes. Among patients with the TT genotype for rs8099917 (associated with a favorable response), the rates of sustained virologic response were higher in patients with a ≥ 3 $\log_{10}$ reduction in serum HCV RNA levels at 4 weeks after starting therapy ($P < 0.0001$). In contrast, among patients with the TG/GG genotype (associated with an unfavorable response), there were no differences in this rate based on the reduction in HCV RNA levels at 4 weeks. Early viral dynamics at 4 weeks after starting therapy retains its predictive value for sustained virologic response in patients with the TT genotype for rs8099917, but not in patients with the TG/GG genotype. Patients who are likely to achieve sustained virologic response despite unfavorable TG/GG genotype cannot be identified based on early viral dynamics during therapy. In contrast, lack of early virologic response at 12 weeks retains a strong predictive value for the failure of sustained virologic response regardless of *IL28B* polymorphisms, which remains useful as a factor to stop therapy. ***J. Med. Virol. 84:61–70, 2012.*** © 2011 Wiley Periodicals, Inc.

KEY WORDS: chronic hepatitis C; early viral dynamics; genetic polymorphisms near the *IL28B* gene; peginterferon; response-guided therapy; ribavirin

INTRODUCTION

The current standard antiviral therapy for patients with chronic hepatitis C is combination therapy with peginterferon (PEG-IFN) and ribavirin [Ghany et al., 2009]. Although this treatment regimen has increased markedly the number of patients with a sustained virologic response, i.e., the eradication of hepatitis C virus (HCV), only 50% of patients infected with HCV genotype 1 achieved a sustained virologic response approximately.

Many investigators have examined factors that predict the treatment outcome of PEG-IFN and ribavirin combination therapy in patients infected with HCV genotype 1. In addition to the baseline factors, the response of HCV during combination therapy, i.e., the changes in serum HCV RNA levels after starting therapy, has been shown to be an important predictor of the treatment outcome [Zeuzem et al., 2001; Buti

Conflict of interest: None.

*Correspondence to: Hidenori Toyoda, MD, PhD, Department of Gastroenterology, Ogaki Municipal Hospital 4-86, Minaminokawa, Ogaki, Gifu, 503-8502, Japan.
E-mail: hmtoyoda@spice.ocn.ne.jp

Accepted 28 September 2011
DOI 10.1002/jmv.22272
Published online in Wiley Online Library (wileyonlinelibrary.com).

et al., 2002; Berg et al., 2003], with the emphasis on "response-guided therapy" [Lee and Ferenci, 2008; Marcellin and Rizzetto, 2008]. Recent reports have emphasized the importance of evaluating the viral dynamics at 4 weeks after starting therapy to predict a sustained virologic response. A rapid virologic response, in which serum HCV RNA is undetectable at 4 weeks after starting therapy, has been the strongest predictive factor of a sustained virologic response reportedly [Martinez-Bauer et al., 2006; Poordad et al., 2008; de Segadas-Soares et al., 2009; Martinot-Peignoux et al., 2009]. In addition, the predictive value of reduced serum HCV RNA levels at 4 weeks after starting therapy has been clarified further, and a ≥ 3 $\log_{10}$ reduction in HCV RNA levels at 4 weeks after starting therapy has high predictive value that a patient will achieve a sustained virologic response as a final outcome, even in the absence of a rapid virologic response [Toyoda et al., 2011].

In contrast, the lack of an early virologic response, defined as either undetectable serum HCV RNA or HCV RNA levels decreased by >2.0 $\log_{10}$ from the pretreatment level at 12 weeks after starting therapy, has been the most important predictor for the failure of a sustained virologic response in patients infected with HCV genotype 1 reportedly [Fried et al., 2002; Davis et al., 2003]. Therefore, treatment may be discontinued in patients without an early virologic response at 12 weeks of treatment, according to the recommendation in the AASLD guidelines [Ghany et al., 2009].

More recently, several studies reported that genetic polymorphisms near the *IL28B* gene (rs8099917, rs12979860) on chromosome 19 affect the virologic response to PEG-IFN and ribavirin combination therapy in patients infected with HCV genotype 1 [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; McCarthy et al., 2010; Rauch et al., 2010]. Furthermore, genetic polymorphisms near the *IL28B* gene are the strongest baseline predictive factor of the final outcome of combination therapy. An additional report showed the effects of genetic polymorphisms near the *IL28B* gene on HCV viral dynamics during PEG-IFN and ribavirin combination therapy [Thompson et al., 2010].

Although early HCV viral dynamics during therapy was shown originally to have a high predictive value for a sustained virologic response in HCV genotype 1-infected patients before genetic polymorphisms near the *IL28B* gene were linked to a therapeutic response, it is not clear whether early viral dynamics retain their predictive value in light of this additional information. The purpose of the present study was to investigate whether response-guided therapy based on viral dynamics at 4 or 12 weeks after initiating therapy retains its ability to predict the final outcome of PEG-IFN and ribavirin combination therapy after accounting for genetic polymorphisms near the *IL28B* gene.

MATERIALS AND METHODS

Patients and Treatment

Between January 2007 and June 2008, a total of 402 patients with chronic hepatitis C received antiviral combination therapy with PEG-IFN and ribavirin for HCV infection at the Ogaki Municipal Hospital or the Nagoya University Hospital. Among these patients, 272 were infected with HCV genotype 1b and had pretreatment HCV RNA levels >5.0 $\log_{10}$ IU/ml based on a quantitative real-time PCR-based method for HCV (HCV COBAS AmpliPrep/COBAS TaqMan System; Roche Molecular Systems, Pleasanton, CA; Lower limit of quantification, 1.7 $\log_{10}$ IU/ml: Lower limit of detection, 1.0 $\log_{10}$ IU/ml) [Colucci et al., 2007; Pittaluga et al., 2008]. This study did not include any patients infected with HCV genotype 1a because this genotype is not found in the general Japanese population.

All patients were given PEG-IFN alpha-2b (Pegintron, Schering-Plough, Tokyo, Japan) weekly and ribavirin (Rebetol, Schering-Plough, Kenilworth, NJ) daily. The PEG-IFN and ribavirin doses were adjusted based on the patient's body weight. Patients weighing $\leq$45 kg were given 60 μg of PEG-IFN alpha-2b once a week, those weighing >45 and $\leq$60 kg were given 80 μg, those weighing >60 and $\leq$75 kg were given 100 μg, those weighing >75 and $\leq$90 kg were given 120 μg, and those weighing >90 kg were given 150 μg. Patients weighing $\leq$60 kg were administered 600 mg of ribavirin per day, those weighing >60 and $\leq$80 kg were given 800 mg per day, and those weighing >80 kg were administered 1000 mg per day. The PEG-IFN and ribavirin doses were modified based on the manufacturer's recommendations. All patients were scheduled to undergo 48 weeks of treatment. The treatment duration was extended up to 72 weeks in some patients. In addition, treatment was discontinued before 48 weeks in some patients who had a low likelihood of achieving an eradication of HCV due to the presence of serum HCV RNA at 24 weeks after starting therapy.

A sustained virologic response was defined as undetectable serum HCV RNA at 24 weeks after ending the therapy. A patient was considered to have relapsed when serum HCV RNA was detectable between the end of treatment and 24 weeks after completing treatment, although serum HCV RNA was undetectable during and at the end of therapy. Patients were considered to have non-response if serum HCV RNA was detectable at 24 weeks after initiating therapy (i.e., null response or partial response according to the American guidelines [Ghany et al., 2009]). Patients were considered to have a rapid virologic response if they had undetectable serum HCV RNA at 4 weeks after starting therapy. An early virologic response was defined as the disappearance or decrease in serum HCV RNA levels by at least 2 $\log_{10}$ at 12 weeks after starting therapy. Patients were considered to have a complete early virologic response if serum HCV RNA was undetectable at 12 weeks after starting therapy and a partial early virologic response if the serum

HCV RNA levels had decreased by at least 2 $\log_{10}$ at 12 weeks after initiating therapy. Patients were considered not to have an early virologic response if their HCV RNA levels did not decrease by more than 2 $\log_{10}$ at 12 weeks compared to the pretreatment levels. Patients were considered to have a slow virologic response if the serum HCV RNA became undetectable between 12 and 24 weeks.

The study protocol was in compliance with the Helsinki Declaration and was approved by the ethics committee of the Ogaki Municipal Hospital and the Nagoya University School of Medicine. Prior to initiating the study, each patient provided written informed consent to use the laboratory data, analyze genetic polymorphisms near the *IL28B* gene, and test stored serum samples.

Assessments of Serum HCV RNA Levels and Genetic Polymorphisms Near the *IL28B* Gene

After a patient provided informed consent, serum samples were obtained at the patient's regular hospital visits, just prior to initiating treatment, every 4 weeks during the treatment period, and during the 24-week follow-up period after treatment. Serum samples were stored at −80°C until further use. The HCV RNA levels were measured using a quantitative real-time PCR-based method for HCV (HCV COBAS AmpliPrep/COBAS TaqMan System).

Genotyping of rs 8099917 polymorphisms near the *IL28B* gene was performed using the TaqMan SNP assay (Applied Biosystems, Foster City, California) according to the manufacturer's guidelines. A predesigned and functionally tested probe was used for rs8099917 (C__11710096_10, Applied Biosystems).

Statistical analyses. Quantitative values are reported as the mean ± SD. In between-group differences were analyzed by the chi-square test. Univariate and multivariate analyses using a logistic regression model were performed to identify factors that predict a sustained virologic response, including age, sex, body weight, serum alanine aminotransferase activity, serum aspartate aminotransferase activity, serum gamma-glutamyl transpeptidase levels, serum alkaline phosphatase values, serum albumin levels, total serum bilirubin values, white blood cell counts, hemoglobin, platelet counts, hepatitis activity grade (A0 and A1 vs. A2 and A3), liver fibrosis grade (F0 and F1 vs. F2 and F3), pretreatment HCV RNA levels ($\geq$6.5 $\log_{10}$ vs. <6.5 $\log_{10}$), reduction in peginterferon dose and ribavirin dose, reduction in HCV RNA levels at 4 weeks after starting therapy ($\geq$3 $\log_{10}$ vs. <3 $\log_{10}$), and the type of an early virologic response. All *P*-values are two-tailed, and $P < 0.05$ was considered significant statistically.

RESULTS

The characteristics of the patients examined in this study are shown in Table I. Liver histology was evaluated according to the METAVIR score [The French METAVIR Cooperative Study Group, 1994]. Although some patients had a reduction in their PEG-IFN and ribavirin doses during therapy, respectively, all patients except for those who discontinued the therapy had more than 80% adhesion to both the PEG-IFN and ribavirin regimens. No patients discontinued the therapy because of adverse effects. The treatment duration was extended up to 72 weeks in 51 of 71 patients (71.8%) who exhibited a slow virologic response. As a final outcome, 118 patients (43.4%) achieved a sustained virologic response, 84 patients (30.9%) relapsed, and the remaining 70 patients (25.7%) had no response.

TABLE I. Characteristics of all Study Patients (n = 272)

Age (years)	56.0 ± 10.9
Sex (female/male)	139 (51.1)/133 (48.9)
Body weight (kg)	57.8 ± 10.5
Alanine aminotransferase (IU/L)	64.6 ± 56.4
Aspartate aminotransferase (IU/L)	53.9 ± 42.7
Gamma-glutamyl transpeptidase (IU)	48.5 ± 43.9
Alkaline phosphatase (IU/L)	267.9 ± 101.3
Albumin (g/dl)	4.04 ± 0.37
Total bilirubin (mg/dl)	0.79 ± 0.30
White blood cell count (/μl)	4892 ± 1333
Hemoglobin (g/dl)	14.0 ± 1.3
Platelet count ($\times 10^3$/μl)	163 ± 51
Liver histology-activity (A0/A1/A2/A3)*	3 (1.2)/136 (55.3)/ 92 (37.4)/15 (6.1)
Liver histology-fibrosis (F0/F1/F2/F3)*	27 (11.0)/114 (46.3)/ 70 (28.5)/35 (14.2)
Pretreatment HCV RNA concentration ($\log_{10}$ IU/ml)	6.35 ± 0.79
Reduction in the peginterferon dose	81 (29.8)
Reduction in the ribavirin dose	130 (47.8)
Final outcomes (sustained virologic response /relapse/ no response)	118 (43.4)/ 84 (30.9)/70 (25.7)

HCV, hepatitis C virus.
Percentages are shown in parentheses.
*Liver biopsy was not performed in 26 patients.

Reduction in Serum HCV RNA Levels at 4 Weeks after Starting Therapy and Treatment Outcome According to Genetic Polymorphisms Near the *IL28B* Gene

An analysis of genetic polymorphisms at rs8099917 near the *IL28B* gene indicated that 207 patients (76.1%) had a TT genotype, 3 patients had a GG genotype (1.1%), and the remaining 62 patients were TG heterozygote (22.8%). Table II shows the comparison of the background characteristics between patients with the favorable TT genotype and those with the unfavorable TG/GG genotype. As reported previously [Abe et al., 2010], gamma-glutamyl transpeptidase level was higher significantly in patients with the TG/GG genotype. As a final outcome, the rate of a sustained virologic response was higher significantly in patients with the TT genotype. Among 207 patients with the TT genotype, serum HCV RNA became undetectable in 19 patients (9.2%) at 4 weeks after starting therapy (a rapid virologic response). In the remaining 188 patients, the decrease in serum HCV RNA levels at 4 weeks after starting therapy ranged from 0.12

TABLE II. Characteristics of Study Patients According to the Genetic Polymorphisms Near the *IL28B* Gene

	Patients with TT genotype of rs8099917 (n = 207)	Patients with TG/GG genotype of rs8099917 (n = 65)	*P*-value
Age (years)	56.5 ± 10.4	54.4 ± 12.4	0.4112
Sex (female/male)	107 (51.7)/100 (48.3)	32 (49.2)/33 (50.8)	0.8384
Body weight (kg)	57.8 ± 10.9	57.8 ± 9.4	0.8361
Alanine aminotransferase (IU/L)	65.1 ± 53.3	62.8 ± 65.6	0.2548
Aspartate aminotransferase (IU/L)	53.6 ± 34.8	54.7 ± 62.0	0.3339
Gamma-glutamyl transpeptidase (IU)	44.2 ± 37.1	62.3 ± 59.0	0.0003
Alkaline phosphatase (IU/L)	263.1 ± 90.3	282.8 ± 129.9	0.3875
Albumin (g/dl)	4.04 ± 0.36	4.05 ± 0.43	0.8020
Total bilirubin (mg/dl)	0.79 ± 0.30	0.76 ± 0.32	0.3010
White blood cell count (/μl)	4826 ± 1333	5100 ± 1320	0.1608
Hemoglobin (g/dl)	13.9 ± 1.3	14.1 ± 1.4	0.3339
Platelet count ($\times 10^3$/μl)	161 ± 49	169 ± 57	0.3871
Liver histology-activity (A0/A1/A2/A3)*	2 (1.1)/98 (52.4)/ 74 (39.6)/13 (6.9)	1 (1.7)/38 (64.4)/ 18 (30.5)/2 (3.4)	0.3241
Liver histology-fibrosis (F0/F1/F2/F3)*	21 (11.2)/83 (44.4)/ 57 (30.5)/26 (13.9)	6 (10.2)/31 (52.5)/ 13 (22.0)/9 (15.3)	0.6401
Pretreatment HCV RNA concentration ($\log_{10}$ IU/ml)	6.37 ± 0.85	6.29 ± 0.55	0.0582
Reduction in the peginterferon dose	61 (29.5)	20 (30.8)	0.9644
Reduction in the ribavirin dose	101 (48.8)	29 (44.6)	0.5565
Final outcomes (sustained virologic response /relapse/ no response)	106 (51.2)/ 69 (33.3)/32 (15.5)	12 (18.4)/15 (23.1)/ 38 (58.5)	<0.0001

HCV, hepatitis C virus.
Percentages are shown in parentheses.
*Liver biopsy was not performed in 26 patients.

$\log_{10}$ to 5.71 $\log_{10}$ (mean, 3.12 $\log_{10}$). The reduction in serum HCV RNA levels was ≥3 $\log_{10}$ in 98 patients (47.3%), <3 $\log_{10}$ and ≥2 $\log_{10}$ in 52 patients (25.1%), <2 $\log_{10}$ and ≥1 $\log_{10}$ in 23 patients (11.1%), and <1 $\log_{10}$ in 15 patients (7.3%). Figure 1A shows the rate of a sustained virologic response according to the reduction in HCV RNA levels at 4 weeks after starting therapy in patients with the TT genotype. The rates were higher significantly in patients who achieved a rapid virologic response or had a ≥3 $\log_{10}$ decrease in

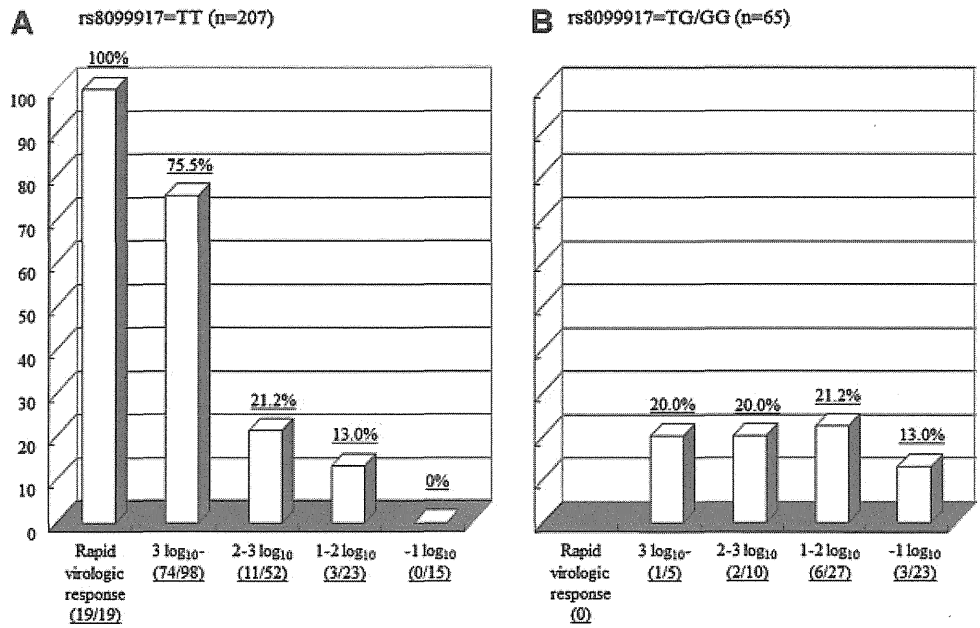

Fig. 1. The rate of sustained virologic responses (%) based on the reduction in serum HCV RNA levels at 4 weeks after starting therapy. **A**: Patients with the TT genotype for rs8099917, (**B**) patients with the TG/GG genotype for rs8099917.

serum HCV RNA levels at 4 weeks compared to those with a <3 $\log_{10}$ decrease in serum HCV RNA levels ($P < 0.0001$). When a 3 $\log_{10}$ decrease in serum HCV RNA levels was defined as the cut-off point, 56.5% of patients were considered to have a ≥3 $\log_{10}$ decrease in serum HCV RNA levels. The sensitivity, specificity, positive predictive value, and negative predictive value for a sustained virologic response were 86.8, 75.2, 78.6, and 84.4%, respectively.

Among the 65 patients who had the TG/GG genotype, no patient achieved a rapid virologic response at 4 weeks after initiating therapy. The decrease in serum HCV RNA levels at 4 weeks after starting therapy ranged from 0.11 $\log_{10}$ to 4.75 $\log_{10}$ (mean, 1.66 $\log_{10}$). The reduction in serum HCV RNA levels at 4 weeks after starting the therapy were smaller in patients with the TG/GG genotype than those with the TT genotype (1.66 ± 1.02 $\log_{10}$ in patients with the TG/GG genotype vs. 3.12 ± 1.37 $\log_{10}$ in patients with TT genotype excluding RVR, $P < 0.0001$). The reduction in serum HCV RNA levels was ≥3 $\log_{10}$ in five patients (7.7%), <3 $\log_{10}$ and ≥2 $\log_{10}$ in 10 patients (15.4%), <2 $\log_{10}$ and ≥1 $\log_{10}$ in 27 patients (41.5%), and <1 $\log_{10}$ in 23 patients (35.4%). Figure 1B shows the rates of a sustained virologic response according to the reduction in HCV RNA levels at 4 weeks after starting therapy in patients with the TG/GG genotype. There were no differences in the rate of a sustained virologic response based on the reduction in HCV RNA levels at 4 weeks after starting therapy; the rate of a sustained virologic response remained at 20% approximately regardless of the reduction in HCV RNA levels in 42 patients with a ≥1 $\log_{10}$ reduction in serum HCV RNA levels.

Association Between an Early Virologic Response at 12 Weeks and Treatment Outcome Based on Genetic Polymorphisms Near the *IL28B* Gene

Figure 2 shows the rate of patients with the TT genotype or TG/GG genotype for rs8099917 who achieved a complete early virologic response, a partial early virologic response, and those who did not achieve early virologic response at 12 weeks after starting therapy based on the reduction in serum HCV RNA level at 4 weeks after initiating therapy. Nearly 75% of patients with the TT genotype whose HCV RNA levels were reduced by ≥3 $\log_{10}$ at 4 weeks after starting the therapy achieved a complete early virologic response. In contrast, 80% of patients with the TG/GG genotype whose HCV RNA levels were reduced by ≥3 $\log_{10}$ at 4 weeks after starting the therapy showed a partial early virologic response. The majority of patients with the TT or TG/GG genotypes achieved a partial early virologic response when their reduction in HCV RNA levels was <3 $\log_{10}$ and ≥2 $\log_{10}$ or <2 $\log_{10}$ and ≥1 $\log_{10}$.

Figure 3 shows the rates of a sustained virologic response according to the type of early virologic response in patients with the TT genotype (Fig. 3A) and TG/GG genotype (Fig. 3B). Among patients with the TT genotype, the rate of sustained virologic response was significantly higher in patients with a complete early virologic response than in those with a partial early virologic response ($P < 0.0001$). In contrast, there was no difference in the rate of a sustained virologic response between patients with a complete early virologic response and those with a partial early virologic response ($P = 0.8917$) among patients with

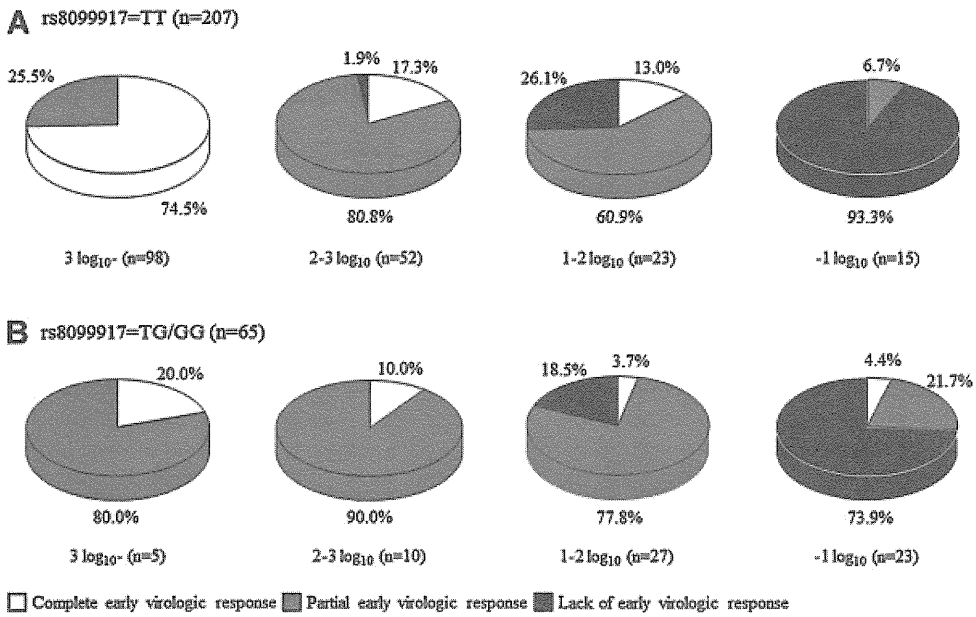

Fig. 2. The association between the virologic responses at 12 weeks after starting therapy and the reduction in serum HCV RNA levels at 4 weeks after starting therapy. A: Patients with the TT genotype for rs8099917, (B) patients with the TG/GG genotype for rs8099917.

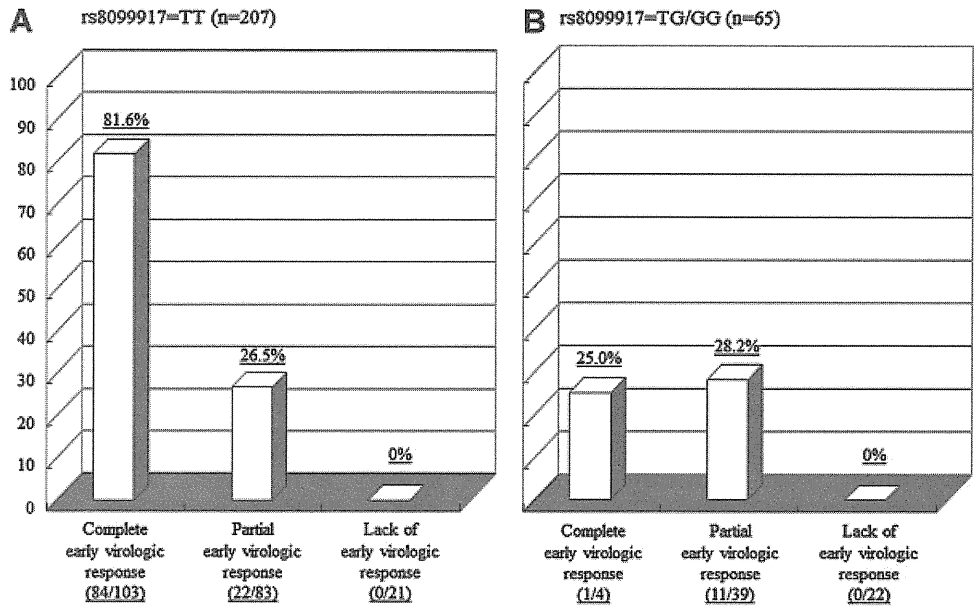

Fig. 3. The rate of sustained virologic responses based on the type of early virologic response. **A**: Patients with the TT genotype for rs8099917, (**B**) patients with the TG/GG genotype for rs8099917.

the TG/GG genotype. None of the patients with the TT genotype or TG/GG genotype who yielded a lack of an early virologic response reached a sustained virologic response.

Univariate and Multivariate Analyses for Factors Associated With a Sustained Virologic Response to Peginterferon and Ribavirin Combination Therapy in Patients With the TT and the TG/GG Genotype for the rs8099917

Univariate and multivariate analyses were conducted for factors associated with a sustained virologic response based on different genetic polymorphisms near the *IL28B* gene. In patients with the TT genotype, the factors that were associated with a sustained virologic response included serum alkaline phosphatase levels, serum albumin, platelet counts, hepatitis activity grade, liver fibrosis grade, reduction in HCV RNA levels at 4 weeks after starting therapy, and a complete early virologic response based on a univariate analysis (Table IIIA). In a multivariate analysis, the serum albumin levels, reduction in HCV RNA levels 4 weeks after starting therapy, and a complete early virologic response were independent factors that were significantly associated with a sustained virologic response (Table IIIB). A reduction in HCV RNA levels 4 weeks after starting therapy was the strongest factor that affected a sustained virologic response. In patients with the TG/GG genotype, the factors that were associated with a sustained virologic response included patient age, platelet counts, and pretreatment HCV RNA levels based on a univariate analysis (Table IIIA). A reduction in the HCV RNA levels at 4 weeks after starting therapy was not associated with a sustained virologic response. In a multivariate analysis, patient age and pretreatment HCV RNA levels were independent factors that were significantly associated with a sustained virologic response (Table IIIC).

Characteristics of Patients who Achieved a Sustained Virologic Response to the Combination Therapy Despite the Unfavorable TG/GG Genotype Near the *IL28B* Gene

Table IV shows the characteristics of 12 patients who achieved a sustained virologic response despite having the unfavorable TG/GG genotype for rs8099917 near the *IL28B* gene. All but one patient was under 60 years old and had liver fibrosis not more than grade 2 (one patient did not undergo a liver biopsy). Except for one patient, the reduction in the serum HCV RNA levels at 4 weeks after starting therapy was less than 3 $\log_{10}$ and all but one patient showed a partial early virologic response at 12 weeks after starting the therapy. In all 11 patients with a partial early virologic response, the serum HCV RNA was undetectable up to 24 weeks after starting the therapy. All but one patient extended the treatment duration from 48 to 72 weeks (two patients discontinued therapy at 60 weeks during the extended treatment period). When the characteristics of patients who achieved a sustained virologic response were compared between those with the unfavorable TG/GG genotype and those with the favorable TT genotype, patients with the TG/GG genotype were younger (41.8 ± 14.4 years vs. 55.1 ± 10.4 years, $P = 0.0023$) and had lower pretreatment HCV RNA levels (5.91 ± 0.44 $\log_{10}$ IU/ml vs. 6.21 ± 1.05 $\log_{10}$ IU/ml, $P = 0.0199$).

TABLE III. Univariate and Multivariate Analyses for Factors Associated With a Sustained Virologic Response to Peginterferon and Ribavirin Combination Therapy in Patients With the TT and the TG/GG Genotype for the rs8099917

(A) Univariate analyses	*P*-value	
	Patients with TT genotype of rs8099917 (n = 207)	Patients with TG/GG genotype of rs8099917 (n = 65)
Age (years)	0.0505	0.0007
Sex (female/male)	0.1830	0.2296
Body weight (kg)	0.6891	0.2456
Alanine aminotransferase (IU/L)	0.7988	0.4032
Aspartate aminotransferase (IU/L)	0.5021	0.1705
Gamma-glutamyl transpeptidase (IU)	0.6340	0.6648
Alkaline phosphatase (IU/L)	0.0315	0.0599
Albumin (g/dl)	0.0002	0.6594
Total bilirubin (mg/dl)	0.2929	0.7130
White blood cell count (/μl)	0.2508	0.5549
Hemoglobin (g/dl)	0.0847	0.2289
Platelet count ($\times 10^3/\mu l$)	0.0454	0.0411
Liver histology-activity (A0–1/A2–3)	0.0445	0.1117
Liver histology-fibrosis (F0–1/F2–3)	0.0002	0.2283
Pretreatment HCV RNA concentration (≥ 6.5 $\log_{10}$ vs. <6.5 $\log_{10}$)	0.5279	0.0379
Reduction in the peginterferon dose	0.4316	0.5563
Reduction in the ribavirin dose	0.1823	0.4272
Reduction in HCV RNA levels at 4 weeks after starting the therapy (≥ 3 $\log_{10}$ vs. < 3 $\log_{10}$)	<0.0001	0.9265
Early virologic response (complete vs. partial)	<0.0001	0.9777
Early virologic response (partial vs. non)	0.8632	0.0686

(B) Multivariate analyses: Patients with TT genotype of rs8099917	*P*-value	Odds ratio (95% confidence interval)
Alkaline phosphatase (IU/L)	0.2617	
Albumin (g/dl)	0.0365	28.287 (1.4107–755.41)
Platelet count ($\times 10^3/\mu l$)	0.2599	
Liver histology-activity (A0–1/A2–3)	0.6678	
Liver histology-fibrosis (F0–1/F2–3)	0.2307	
Reduction in HCV RNA levels at 4 weeks after starting the therapy (≥ 3 $\log_{10}$ vs. <3 $\log_{10}$)	<0.0001	16.029 (6.8593–40.406)
Early virologic response (complete vs. partial)	0.0224	0.3685 (0.1557–0.8749)

(C) Multivariate analyses: Patients with TG/GG genotype of rs8099917	*P*-value	Odds ratio (95% confidence interval)
Age (years)	0.0022	0.0034 (0.0000–0.0840)
Platelet count ($\times 10^3/\mu l$)	0.3344	
Pretreatment HCV RNA concentration (≥ 6.5 $\log_{10}$ vs. <6.5 $\log_{10}$)	0.0304	0.0548 (0.0020–0.4950)

HCV, hepatitis C virus.

DISCUSSION

Several previous studies reported that patients who achieved a rapid virologic response, in which serum HCV RNA become undetectable at 4 weeks after starting therapy, had a high likelihood of achieving a sustained virologic response [Martinez-Bauer et al., 2006; Poordad et al., 2008; de Segadas-Soares et al., 2009; Martinot-Peignoux et al., 2009]. In addition, several recent studies reported the predictive value of the degree of reduction in serum HCV RNA levels at 4 weeks after starting therapy [Yu et al., 2007; Huang et al., 2010; Toyoda et al., 2011]. Therefore, the viral dynamics of HCV at 4 as well as 12 weeks after starting therapy is important for response-guided therapy.

Genetic polymorphisms near the *IL28B* gene have emerged as the strongest predictive factor of a sustained virologic response in patients infected with HCV genotype 1 [Hayes et al., 2011; Kurosaki et al., 2011]. In addition, Thompson et al. [2010 reported that genetic polymorphisms near the *IL28B* gene were associated strongly with early viral dynamics during PEG-IFN and ribavirin combination therapy. These findings raised an important issue of whether response-guided therapy, based on the reduction in serum HCV RNA levels at 4 or 12 weeks after starting

TABLE IV. Patients who Achieved a Sustained Virologic Response Despite the TG/GG Genotype for the rs8099917

	Age (years)	Sex	Liver histology	Pretreatment HCV RNA level (log_{10} IU/ml)	HCV RNA reduction at 4 weeks	Response at 12 weeks	HCV RNA became undetectable (weeks)	Treatment duration (weeks)
1.	31	Female	A1/F1	6.13	2.19	partial EVR	20	48
2.	55	Male	A1/F1	5.80	1.77	partial EVR	16	72
3.	57	Female	A1/F1	5.58	3.01	partial EVR	16	72
4.	57	Female	A1/F1	6.21	1.81	partial EVR	20	72
5.	62	Male	N.D.	6.23	1.13	partial EVR	24	72
6.	21	Male	A1/F2	6.04	1.83	partial EVR	24	72
7.	42	Male	A1/F1	6.27	0.57	partial EVR	24	72
8.	29	Female	A1/F2	5.83	1.83	partial EVR	20	60
9.	52	Male	A1/F0	5.91	2.12	complete EVR	12	48
10.	40	Male	A2/F1	5.84	1.34	partial EVR	20	72
11.	27	Male	N.D.	5.63	0.42	partial EVR	24	72
12.	28	Male	A1/F0	6.59	0.76	partial EVR	20	60

N.D., not done; HCV, hepatitis C virus; EVR, early virologic response.

therapy, retains a predictive value when considering genetic polymorphisms near the *IL28B* gene.

In the present study, the predictive value of the decrease in serum HCV RNA levels was evaluated at 4 and 12 weeks after starting therapy in Japanese patients infected with HCV genotype 1b based on genetic polymorphisms near the *IL28B* gene. Consistent with previous reports, patients with the TG/GG genotype for rs8099917 had a smaller reduction in serum HCV RNA levels at 4 weeks after starting treatment ($P < 0.0001$), which indicates an unfavorable response to the combination therapy. Patients with the TT genotype for rs8099917, which is associated with a favorable response to the combination therapy, exhibited a significant difference in the rate of a sustained virologic response based on the reduction in serum HCV RNA levels at 4 weeks after initiating the therapy. Patients with a rapid virologic response or with a ≥ 3 log_{10} reduction in HCV RNA levels had a higher likelihood of achieving a sustained virologic response.

In contrast, these factors did not have any predictive value in patients with the TG/GG genotype. Only 18.5% of patients achieved a sustained virologic response (12 of 65 patients), and it was difficult to identify these patients based on the reduction in HCV RNA levels at 4 weeks or the type of an early virologic response at 12 weeks after starting therapy. Patients who achieved a sustained virologic response, despite the TG/GG genotype for rs8099917, were identified among those with a <2 log_{10} and ≥ 1 log_{10} or even <1 log_{10} reduction in HCV RNA levels at 4 weeks after starting therapy. Interestingly and paradoxically, the possibility of a sustained virologic response can be expected in patients with a <1 log_{10} reduction in HCV RNA levels at 4 weeks after starting therapy only when they have the unfavorable TG/GG genotype.

In the evaluation at 12 weeks after starting therapy, patients with the TT genotype who achieved a complete early virologic response had a higher rate of a sustained virologic response significantly than patients who achieved a partial early virologic response, whereas this difference was not found in patients with the TG/GG genotype. No patients who failed to achieve an early virologic response achieved a sustained virologic response regardless of the genetic polymorphisms near the *IL28B* gene. Thus, the lack of an early virologic response retained a strong predictive value for the failure of achieving a sustained virologic response. This result supports the recommendation in the AASLD guidelines, in which treatment may be discontinued in patients without an early virologic response at 12 weeks of treatment.

The characteristics of patients who achieved a sustained virologic response despite the unfavorable TG/GG genotype were younger in age and lower pretreatment HCV RNA levels. Most patients with the TG/GG genotype who achieved a sustained virologic response showed a partial early virologic response and extended the treatment duration. It was difficult to identify these patients according to viral dynamics at 4 or 12 weeks after starting therapy.

There are several limitations in this study. Some patients with a slow virologic response did not have their treatment period extended from 48 to 72 weeks. This is because the effectiveness of a 72-week combination therapy regimen in patients with HCV genotype 1 with a slow virologic response [Berg et al., 2006; Pearlman et al., 2007] had not been established in Japan in the earlier part of this study. This fact might have influenced the treatment outcome especially in patients with the unfavorable TG/GG genotype. Another limitation is a smaller sample size of patients with the TG/GG genotype in comparison to that of patients with the TT genotype. This sample size could have caused the lack of statistical significance in the rate of a sustained virologic response according to the reduction in HCV RNA levels at 4 weeks after starting therapy or according to the type of an early virologic response in patients with the TG/GG genotype. In addition, the data were based on Japanese patients infected with HCV genotype 1b. Therefore, these results should be confirmed in other ethnicities and patients infected with HCV genotype 1a.