

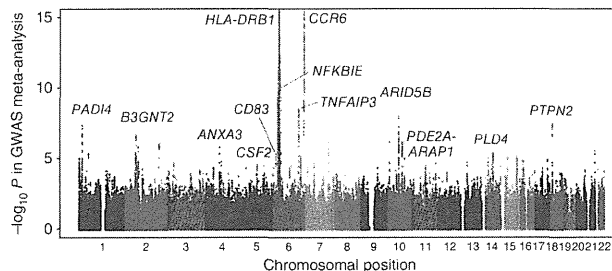
## Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population

Yukinori Okada<sup>1-3,40</sup>, Chikashi Terao<sup>4,5,40</sup>, Katsunori Ikari<sup>6,40</sup>, Yuta Kochi<sup>1,2,40</sup>, Koichiro Ohmura<sup>5</sup>, Akari Suzuki<sup>1</sup>, Takahisa Kawaguchi<sup>4</sup>, Eli A Stahl<sup>7,8</sup>, Fina A S Kurreeman<sup>7-9</sup>, Nao Nishida<sup>10</sup>, Hiroko Ohmiya<sup>3</sup>, Keiko Myouzen<sup>1</sup>, Meiko Takahashi<sup>4</sup>, Tetsuji Sawada<sup>11</sup>, Yuichi Nishioka<sup>12</sup>, Masao Yukioka<sup>13</sup>, Tsukasa Matsubara<sup>14</sup>, Shigeyuki Wakitani<sup>15</sup>, Ryota Teshima<sup>16</sup>, Shigeto Tohma<sup>17</sup>, Kiyoshi Takasugi<sup>18</sup>, Kota Shimada<sup>17</sup>, Akira Murasawa<sup>19</sup>, Shigeru Honjo<sup>20</sup>, Keitaro Matsuo<sup>21</sup>, Hideo Tanaka<sup>21</sup>, Kazuo Tajima<sup>22</sup>, Taku Suzuki<sup>6,23</sup>, Takuji Iwamoto<sup>6,23</sup>, Yoshiya Kawamura<sup>24</sup>, Hisashi Tanii<sup>25</sup>, Yuji Okazaki<sup>26</sup>, Tsukasa Sasaki<sup>27</sup>, Peter K Gregersen<sup>28</sup>, Leonid Padyukov<sup>29</sup>, Jane Worthington<sup>30</sup>, Katherine A Siminovitch<sup>31</sup>, Mark Lathrop<sup>32,33</sup>, Atsuo Taniguchi<sup>6</sup>, Atsushi Takahashi<sup>3</sup>, Katsushi Tokunaga<sup>10</sup>, Michiaki Kubo<sup>34</sup>, Yusuke Nakamura<sup>35</sup>, Naoyuki Kamatani<sup>36</sup>, Tsuneyo Mimori<sup>5</sup>, Robert M Plenge<sup>7,8</sup>, Hisashi Yamanaka<sup>6</sup>, Shigeki Momohara<sup>6,41</sup>, Ryo Yamada<sup>37,41</sup>, Fumihiko Matsuda<sup>4,38,39,41</sup> & Kazuhiko Yamamoto<sup>1,2,41</sup>

Rheumatoid arthritis is a common autoimmune disease characterized by chronic inflammation. We report a meta-analysis of genome-wide association studies (GWAS) in a Japanese population including 4,074 individuals with rheumatoid arthritis (cases) and 16,891 controls, followed by a replication in 5,277 rheumatoid arthritis cases and 21,684 controls. Our study identified nine loci newly associated with rheumatoid arthritis at a threshold of  $P < 5.0 \times 10^{-8}$ , including *B3GNT2*, *ANXA3*, *CSF2*, *CD83*, *NFKBIE*, *ARID5B*, *PDE2A-ARAP1*, *PLD4* and *PTPN2*. *ANXA3* was also associated with susceptibility to systemic lupus erythematosus ( $P = 0.0040$ ), and *B3GNT2* and *ARID5B* were associated with Graves' disease ( $P = 3.5 \times 10^{-4}$  and  $2.9 \times 10^{-4}$ , respectively). We conducted a multi-ancestry comparative analysis with a previous meta-analysis in individuals of European descent (5,539 rheumatoid arthritis cases and 20,169 controls). This provided evidence of shared genetic risks of rheumatoid arthritis between the populations.

Rheumatoid arthritis is a complex autoimmune disease characterized by inflammation and the destruction of synovial joints and affects up to 1% of the population worldwide. To date, more than 35 rheumatoid arthritis susceptibility loci, including *HLA-DRB1*, *PTPN22*, *PADI4*, *STAT4*, *TNFAIP3* and *CCR6*, among others, have been identified by GWAS in multiple populations<sup>1-12</sup> and by several meta-analyses of the original GWAS<sup>13-16</sup>. In particular, each meta-analysis of these GWAS uncovered a number of loci that were not identified in the single GWAS, leading to recognition of the enormous power of the meta-analysis approach for detecting causal genes in disease. However, these previous meta-analyses have been performed solely in European populations<sup>13-16</sup> and not in

Asian ones. As multi-ancestry studies on validated rheumatoid arthritis susceptibility loci showed the existence of both population-specific and shared genetic components of rheumatoid arthritis<sup>10,17</sup>, additional studies in Asian populations might provide useful insight into the underlying genetic architecture of rheumatoid arthritis, which would otherwise be difficult to capture using the studies in a single population. Here, we report a meta-analysis of GWAS and a replication study for rheumatoid arthritis in a Japanese population that was conducted by the Genetics and Allied research in Rheumatic diseases NETworking (GARNET) consortium<sup>10,12</sup>. We subsequently performed a multi-ancestry comparative analysis that incorporated results from a previously conducted meta-analysis of individuals of European ancestry<sup>15</sup>.



**Figure 1** Manhattan plots of the GWAS meta-analysis for rheumatoid arthritis in the Japanese population. The genetic loci that satisfied the genome-wide significance threshold of  $P < 5.0 \times 10^{-8}$  (gray line) in the meta-analysis or in the combined study of the meta-analysis and the replication study are presented. The y axis shows the  $-\log_{10} P$  values of the SNPs in the meta-analysis. The SNPs for which the  $P$  values were smaller than  $1.0 \times 10^{-15}$  are indicated at the upper limit of the plot.

A full list of author affiliations appears at the end of the paper.

Received 24 October 2011; accepted 1 March 2012; published online 25 March 2012; doi:10.1038/ng.2231

Table 1 Results of the GWAS meta-analysis and the replication studies for rheumatoid arthritis

| rsID <sup>a</sup>  | Chr. | Position (bp) | Cytoband | Gene(s)     | Associations in Japanese |       |       |                  |                          |                   |                          |                  |                          |       | Associations in Europeans <sup>b</sup> |                  |                          |         |                          |
|--|------|---------------|----------|-------------|--------------------------|-------|-------|------------------|--------------------------|-------------------|--------------------------|------------------|--------------------------|-------|--|------------------|--------------------------|---------|--------------------------|
|  |      |               |          |             | GWAS meta-analysis       |       |       |                  |                          | Replication study |                          |                  |                          |       | GWAS meta-analysis                     |                  |                          |         |                          |
|  |      |               |          |             | Allele                   | 1/2   | RA    | Control          | OR (95% CI) <sup>b</sup> | P                 | OR (95% CI) <sup>b</sup> | P                | OR (95% CI) <sup>b</sup> | P     | OR (95% CI) <sup>b</sup>               | P                | RA                       | Control | OR (95% CI) <sup>b</sup> |
| SNPs with significant associations ( $P < 5.0 \times 10^{-8}$ in the combined study)                         |      |               |          |             |                          |       |       |                  |                          |                   |                          |                  |                          |       |  |                  |                          |         |                          |
| rs11900673   | 2    | 62306165      | 2p15     | B3GNT2      | T/C                      | 0.31  | 0.28  | 1.15 (1.08–1.21) | $3.5 \times 10^{-6}$     | 1.09 (1.04–1.14)  | $6.0 \times 10^{-4}$     | 1.11 (1.07–1.15) | $1.1 \times 10^{-8}$     | 0.13  | 0.13                                   | 1.05 (0.98–1.13) | 0.17                     |         |                          |
| rs2867461  | 4    | 79732239      | 4q21     | ANXA3       | A/G                      | 0.46  | 0.44  | 1.13 (1.08–1.19) | $4.7 \times 10^{-6}$     | 1.12 (1.08–1.17)  | $1.2 \times 10^{-7}$     | 1.13 (1.09–1.17) | $1.2 \times 10^{-12}$    | 0.37  | 0.37                                   | 0.98 (0.92–1.04) | 0.52                     |         |                          |
| rs657075   | 5    | 131458017     | 5q31     | CSF2        | A/G                      | 0.38  | 0.36  | 1.12 (1.06–1.18) | $3.2 \times 10^{-5}$     | 1.11 (1.06–1.16)  | $3.8 \times 10^{-6}$     | 1.12 (1.08–1.15) | $2.8 \times 10^{-10}$    | 0.10  | 0.10                                   | 1.04 (0.95–1.13) | 0.37                     |         |                          |
| rs12529514   | 6    | 14204637      | 6p23     | CD83        | C/T                      | 0.16  | 0.14  | 1.19 (1.10–1.27) | $6.8 \times 10^{-6}$     | 1.11 (1.05–1.18)  | $6.0 \times 10^{-4}$     | 1.14 (1.09–1.19) | $2.0 \times 10^{-8}$     | 0.055 | 0.053                                  | 1.11 (0.99–1.24) | 0.074                    |         |                          |
| rs22393434   | 6    | 44340898      | 6p21.1   | NFKBIE      | G/A                      | 0.24  | 0.21  | 1.23 (1.16–1.31) | $9.2 \times 10^{-11}$    | 1.17 (1.11–1.23)  | $2.2 \times 10^{-9}$     | 1.19 (1.15–1.24) | $5.8 \times 10^{-19}$    | 0.059 | 0.040                                  | 1.57 (1.11–2.21) | 0.0099                   |         |                          |
| rs10821944   | 10   | 63455095      | 10q21    | ARID5B      | G/T                      | 0.39  | 0.36  | 1.17 (1.11–1.23) | $1.0 \times 10^{-8}$     | 1.15 (1.10–1.20)  | $3.0 \times 10^{-10}$    | 1.16 (1.12–1.20) | $5.5 \times 10^{-18}$    | 0.29  | 0.26                                   | 1.11 (1.05–1.17) | $1.9 \times 10^{-4}$     |         |                          |
| rs781913   | 11   | 72051144      | 11q13    | PDE2A-ARAP1 | T/G                      | 0.71  | 0.69  | 1.11 (1.05–1.17) | $3.2 \times 10^{-4}$     | 1.13 (1.08–1.18)  | $6.7 \times 10^{-7}$     | 1.12 (1.08–1.16) | $5.8 \times 10^{-10}$    | 0.45  | 0.43                                   | 1.04 (0.99–1.09) | 0.13                     |         |                          |
| rs2841277  | 14   | 104462050     | 14q32    | PLD4        | T/C                      | 0.72  | 0.69  | 1.11 (1.05–1.18) | $2.8 \times 10^{-4}$     | 1.18 (1.13–1.24)  | $7.0 \times 10^{-12}$    | 1.15 (1.11–1.19) | $1.9 \times 10^{-14}$    | 0.47  | 0.46                                   | 1.02 (0.96–1.09) | 0.54                     |         |                          |
| rs2847297  | 18   | 12787694      | 18p11    | PTPN2       | G/A                      | 0.37  | 0.33  | 1.16 (1.11–1.23) | $3.5 \times 10^{-8}$     | 1.06 (1.01–1.11)  | 0.013                    | 1.10 (1.07–1.14) | $2.2 \times 10^{-8}$     | 0.36  | 0.34                                   | 1.10 (1.05–1.15) | $9.2 \times 10^{-5}$     |         |                          |
| SNPs with suggestive associations ( $5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-5}$ in the combined study)  |      |               |          |             |                          |       |       |                  |                          |                   |                          |                  |                          |       |  |                  |                          |         |                          |
| rs4937362  | 11   | 127997949     | 11q24    | ETS1-FL1    | T/C                      | 0.71  | 0.68  | 1.13 (1.07–1.19) | $2.0 \times 10^{-5}$     | 1.07 (1.02–1.12)  | 0.0061                   | 1.09 (1.06–1.13) | $7.5 \times 10^{-7}$     | 0.46  | 0.44                                   | 1.06 (1.01–1.11) | 0.015                    |         |                          |
| rs3783637  | 14   | 54417868      | 14q22    | GCH1        | C/T                      | 0.76  | 0.74  | 1.13 (1.07–1.20) | $6.5 \times 10^{-5}$     | 1.07 (1.02–1.13)  | 0.0062                   | 1.10 (1.06–1.14) | $2.0 \times 10^{-6}$     | 0.88  | 0.88                                   | 0.99 (0.88–1.11) | 0.87                     |         |                          |
| rs1957895  | 14   | 60978085      | 14q23    | PRKCH       | G/T                      | 0.40  | 0.39  | 1.12 (1.06–1.18) | $4.1 \times 10^{-5}$     | 1.07 (1.02–1.12)  | 0.0022                   | 1.09 (1.05–1.13) | $3.6 \times 10^{-7}$     | 0.093 | 0.089                                  | 1.01 (0.95–1.07) | 0.73                     |         |                          |
| rs6496667  | 15   | 88694672      | 15q26    | ZNF774      | A/C                      | 0.38  | 0.35  | 1.13 (1.07–1.19) | $4.7 \times 10^{-5}$     | 1.07 (1.02–1.11)  | 0.0050                   | 1.09 (1.05–1.13) | $1.4 \times 10^{-6}$     | 0.21  | 0.20                                   | 1.07 (1.01–1.13) | 0.031                    |         |                          |
| rs7404928  | 16   | 23796341      | 16p12    | PRKCB1      | T/C                      | 0.65  | 0.62  | 1.13 (1.07–1.19) | $1.5 \times 10^{-9}$     | 1.05 (1.01–1.10)  | 0.026                    | 1.08 (1.05–1.12) | $4.0 \times 10^{-6}$     | 0.75  | 0.75                                   | 1.01 (0.94–1.09) | 0.79                     |         |                          |
| rs2280381  | 16   | 84576134      | 16q24    | IRF8        | T/C                      | 0.86  | 0.84  | 1.16 (1.08–1.25) | $1.0 \times 10^{-4}$     | 1.09 (1.03–1.15)  | 0.0049                   | 1.12 (1.07–1.17) | $2.4 \times 10^{-6}$     | 0.62  | 0.60                                   | 1.05 (0.99–1.11) | 0.081                    |         |                          |
| SNPs in previously reported rheumatoid arthritis susceptibility loci ( $P < 5.0 \times 10^{-8}$ in the GWAS) |      |               |          |             |                          |       |       |                  |                          |                   |                          |                  |                          |       |  |                  |                          |         |                          |
| rs766449   | 1    | 17547439      | 1p36     | PADI4       | T/C                      | 0.44  | 0.40  | 1.17 (1.11–1.24) | $4.6 \times 10^{-8}$     | -                 | -                        | -                | -                        | 0.38  | 0.37                                   | 1.09 (1.03–1.05) | 0.0022                   |         |                          |
| rs2157337  | 6    | 32609122      | 6p21.3   | HLA-DRB1    | C/T                      | 0.59  | 0.44  | 1.99 (1.88–2.11) | $2.6 \times 10^{-118}$   | -                 | -                        | -                | -                        | 0.69  | 0.46                                   | 2.50 (2.39–2.62) | $< 1.0 \times 10^{-840}$ |         |                          |
| rs6932056  | 6    | 138284130     | 6q23     | TNFAIP3     | C/T                      | 0.092 | 0.073 | 1.35 (1.23–1.49) | $3.2 \times 10^{-9}$     | -                 | -                        | -                | -                        | 0.044 | 0.034                                  | 1.41 (1.24–1.60) | $1.3 \times 10^{-7}$     |         |                          |
| rs15171878   | 6    | 167460832     | 6q27     | CCR6        | C/T                      | 0.54  | 0.48  | 1.31 (1.24–1.39) | $3.2 \times 10^{-19}$    | -                 | -                        | -                | -                        | 0.47  | 0.43                                   | 1.13 (1.08–1.19) | $5.9 \times 10^{-7}$     |         |                          |

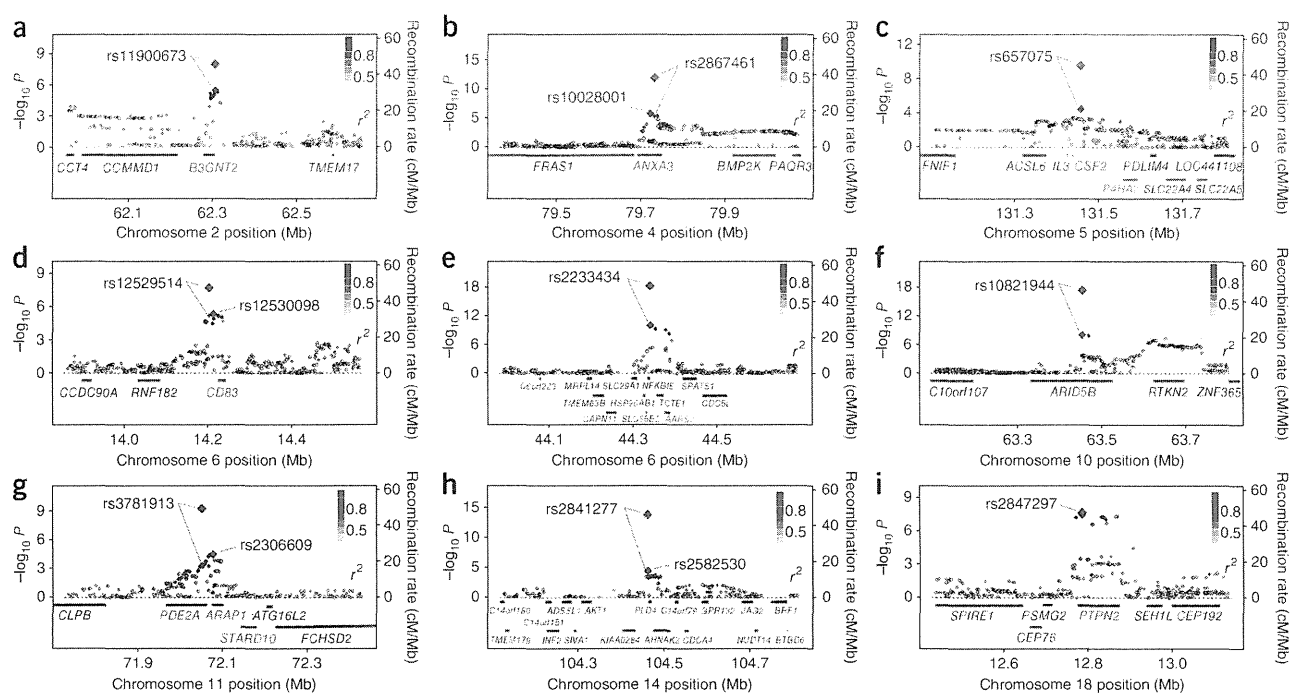
Chr., chromosome; Freq., frequency; RA, rheumatoid arthritis; OR, odds ratio; CI, confidence interval.

<sup>a</sup>SNPs with  $P < 5.0 \times 10^{-8}$  in the combined study of the GWAS meta-analysis and the replication study or SNPs with  $P < 5.0 \times 10^{-5}$  in the GWAS meta-analysis are annotated according to forward strand and NCBI Build 36.3. Full results of the replication study are provided in Supplementary Table 3. <sup>b</sup>Odds ratio of allele 1. <sup>c</sup>Associations in the previous meta-analysis in European populations<sup>15</sup>.

The meta-analysis included 4,074 rheumatoid arthritis cases (with 81.4% and 80.4% of the subjects being positive for antibody to cyclic citrullinated peptide (anti-CCP) and rheumatoid factor, respectively) and 16,891 controls from three GWAS of Japanese subjects (from the BioBank Japan Project<sup>10,18</sup>, Kyoto University<sup>12</sup> and the Institute of Rheumatology Rheumatoid Arthritis (IORRA)<sup>19</sup>; **Supplementary Table 1**). After the application of stringent quality control criteria, including principal-component analysis (PCA; **Supplementary Fig. 1**) for each GWAS, the meta-analysis was conducted by evaluating ~2.0 million autosomal SNPs with minor allele frequencies (MAFs)  $\geq 0.01$ , which were obtained through whole-genome imputation of genotypes on the basis of the HapMap Phase 2 East Asian panels (Japanese in Tokyo (JPT) and Han Chinese in Beijing (CHB)). The inflation factor of the test statistics in the meta-analysis  $\lambda_{GC}$  was as low as 1.036, suggesting no substantial effects of population structure (**Supplementary Table 2**). The quantile-quantile plot of  $P$  values showed a marked discrepancy in the values in its tail from those anticipated under the null hypothesis that there is no association—even after removal of the SNPs located in the human leukocyte antigen (HLA) region, the major rheumatoid arthritis susceptibility locus—thereby showing the presence of significant associations in the meta-analysis (**Supplementary Fig. 2**).

We identified seven loci in the current meta-analysis that satisfied the genome-wide significance threshold of  $P < 5.0 \times 10^{-8}$ . These included previously known rheumatoid arthritis susceptibility loci, such as *PADI4* at 1p36, *HLA-DRB1* at 6p21.3, *TNFAIP3* at 6q23 and *CCR6* at 6q27 (refs. 1,3,6,10,15) (the smallest  $P = 2.6 \times 10^{-118}$  was found at the *HLA-DRB1* locus; **Fig. 1** and **Table 1**). To our knowledge, the other three loci identified, *NFKBIE* at 6p21.1, *ARID5B* at 10q21 and *PTPN2* at 18p11, are newly associated ( $P = 9.2 \times 10^{-11}$ ,  $1.0 \times 10^{-8}$  and  $3.5 \times 10^{-8}$ , respectively).

To validate the associations identified in the meta-analysis, we conducted a replication study of two independent Japanese rheumatoid arthritis case-control cohorts (cohort 1: 3,830 rheumatoid arthritis cases and 17,920 controls, cohort 2: 1,447 rheumatoid arthritis cases and 3,764 controls; **Supplementary Table 1**). To increase the number of subjects and enhance statistical power, genotype data obtained from other GWAS projects conducted for non-autoimmune diseases in Japanese using Illumina platforms were used for the replication control panels. For each of the 46 loci that exhibited  $P < 5.0 \times 10^{-4}$  in



**Figure 2** Regional plots of the loci newly associated with rheumatoid arthritis at the genome-wide significance threshold of  $P < 5.0 \times 10^{-8}$  in the combined study of the meta-analysis and the replication study. (a–i) Regional plots are shown at *B3GNT2* (a), *ANXA3* (b), *CSF2* (c), *CD83* (d), *NFKBIE* (e), *ARID5B* (f), *PDE2A-ARAP1* (g), *PLD4* (h) and *PTPN2* (i). Diamonds represent the  $-\log_{10} P$  values of the SNPs, and the red diamonds represent the  $-\log_{10} P$  values of the SNPs in the meta-analysis. Red color for the smaller circles represents the  $r^2$  value with the most significantly associated SNP (larger red circle). The purple circle represents the  $P$  value in the combined study. The blue line shows the recombination rates given by the HapMap Phase 2 east Asian populations (release 22). RefSeq genes at the loci are indicated below. Genes nearest to the marker SNPs at the loci are colored blue (Supplementary Note), and genes implicated in eQTL analysis are colored red (Supplementary Table 4). At 11q13, two genes (*PDE2A* and *ARAP1*) that are nearest to the SNP selected for the replication study and the most significant SNP in the meta-analysis are highlighted. The plots were drawn using SNP Annotation and Proxy Search (SNAP) version 2.2.

the meta-analysis and had not been reported as rheumatoid arthritis susceptibility loci<sup>1–16</sup>, we selected a marker SNP for the replication study (Online Methods and Supplementary Table 3).

In the combined analyses of the meta-analysis and the replication study, including a total of 9,351 rheumatoid arthritis cases and 38,575 controls, we identified six newly associated loci, in addition to the *NFKBIE*, *ARID5B* and *PTPN2* loci, that satisfied the significance threshold of  $P < 5.0 \times 10^{-8}$ , including *B3GNT2* at 2p15, *ANXA3* at 4q21, *CSF2* at 5q31, *CD83* at 6p23, *PDE2A-ARAP1* at 11q13 and *PLD4* at 14q32 (Figs. 1 and 2 and Table 1). Of these loci, *NFKBIE* had the smallest  $P$  value ( $5.8 \times 10^{-19}$ ). Although association with rheumatoid arthritis has been described for the *CSF2* and *PTPN2* loci<sup>11,15,16,20,21</sup>, ours is the first report to our knowledge validating these associations with a threshold of  $P < 5.0 \times 10^{-8}$ . Suggestive associations were also observed in *ETS1-FLI1* at 11q24, *GCH1* at 14q22, *PRKCH* at 14q23, *ZNF774* at 15q26, *PRKCB1* at 16p12 and *IRF8* at 16q24 ( $5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-6}$ ). A summary of the genes in the newly associated loci and the results of *cis* expression quantitative trait locus (*cis* eQTL) analysis of the marker SNPs are provided (Supplementary Table 4 and Supplementary Note).

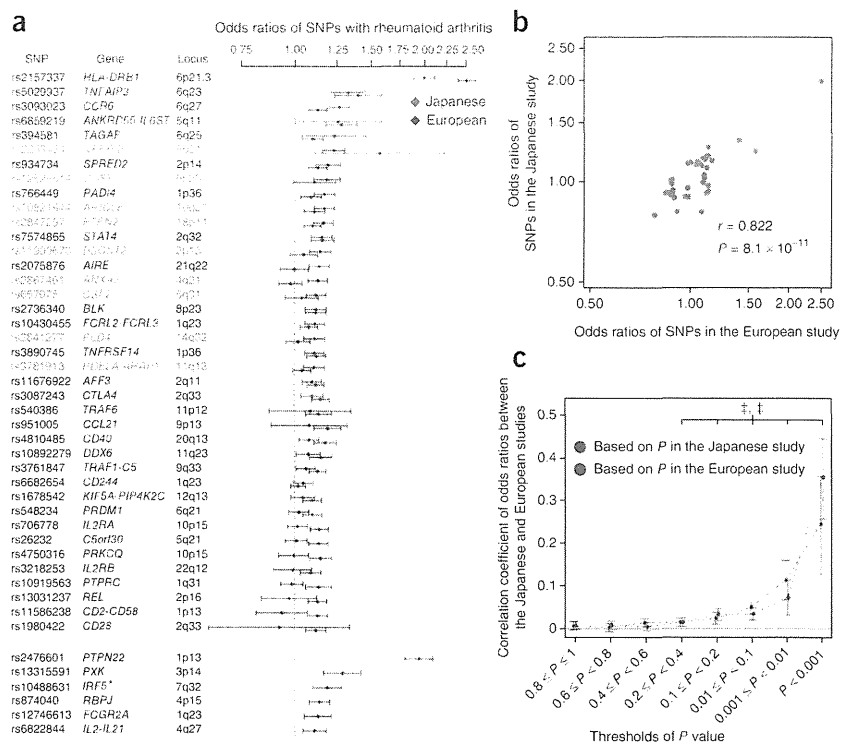
Previous studies have reported associations of rheumatoid arthritis susceptibility loci with other autoimmune diseases<sup>4,10,15,16</sup>. Therefore, we assessed the association of these newly identified susceptibility loci with systemic lupus erythematosus (SLE) by examining the results of an SLE GWAS in the Japanese population (891 cases and 3,384 controls)<sup>22</sup> and in Graves' disease by genotyping 1,783 cases<sup>10</sup> (the controls from the SLE analysis were used for testing for Graves'

disease). We observed significant associations of the *ANXA3* locus with SLE and of the *B3GNT2* and *ARID5B* loci with Graves' disease, which showed the same directional effects of the alleles as in rheumatoid arthritis ( $P < 0.05/9 = 0.0056$ , Bonferroni correction of the number of loci; Supplementary Table 5). It should be noted that relatively small sample sizes in the SLE and Graves' disease cohorts might yield limited statistical power, and further evaluations enrolling larger numbers of subjects would be desirable.

To highlight genetic backgrounds of rheumatoid arthritis that are common and divergent in different ancestry groups, we conducted a multi-ancestry comparative analysis of the present study in Japanese and a previous GWAS meta-analysis in Europeans that included 5,539 rheumatoid arthritis cases and 20,169 controls<sup>15</sup> (Fig. 3a–c). First, we compared associations in the reported<sup>1–16</sup> or newly identified rheumatoid arthritis susceptibility loci (Fig. 3a and Supplementary Table 6). Of the 46 rheumatoid arthritis risk variants evaluated, 6 were monomorphic in Japanese, and all were polymorphic in Europeans. We observed significant associations at 22 loci in Japanese and at 36 loci in Europeans (false discovery rate (FDR)  $< 0.05$ ,  $P < 0.0030$ ), with 14 loci being shared between the populations. Of the newly associated rheumatoid arthritis susceptibility loci identified in our Japanese meta-analysis, significant associations were also observed in the European meta-analysis at the *ARID5B* and *PTPN2* loci ( $P = 1.9 \times 10^{-4}$  and  $9.2 \times 10^{-5}$ , respectively; Table 1). Significant positive correlation of odds ratios was observed between the studies ( $r = 0.822$ ,  $P = 8.1 \times 10^{-11}$ ; Fig. 3b), suggesting that a substantial proportion of genetic factors are shared between

## LETTERS

**Figure 3** Overlap of the associations with rheumatoid arthritis between Japanese and European populations. (a) Forest plots of SNPs in the rheumatoid arthritis susceptibility loci (Supplementary Table 6). We selected the genetic loci that have been validated to be associated with rheumatoid arthritis susceptibility by showing associations in the reports of multiple cohorts or satisfying the genome-wide significant threshold ( $P < 5.0 \times 10^{-8}$ ) in previous studies, including in the meta-analysis and replication phases<sup>1–16</sup>. For each of the loci, the most significant SNP among those reported in the previous or present study were selected<sup>1–16</sup>. SNPs in the newly identified rheumatoid arthritis susceptibility loci are colored green. Odds ratios and 95% confidence interval (CI) values are based on rheumatoid arthritis risk alleles, and the SNPs are ordered according to the odds ratios in the Japanese study. Several SNPs were monomorphic in the Japanese population. The odds ratios of these SNPs in the European study are presented below. The asterisk indicates that an association of another variant at the *IRF5* locus was reported in the Japanese population<sup>24</sup>. (b) Correlation of the odds ratios of the SNPs in the validated rheumatoid arthritis susceptibility loci between the two populations. SNPs that were polymorphic in both populations were used; odds ratios were based on the minor allele in the Japanese population. (c) Correlation of the odds ratios of the genome-wide SNPs, excluding the rheumatoid arthritis susceptibility loci. Correlations were evaluated for sets of SNPs stratified by the thresholds based on the meta-analysis  $P$  values in each population after pruning of the SNPs by LD ( $r^2 < 0.3$ ). Correlation coefficient and 95% CI are indicated on the y axis. Significant correlation of the odds ratios was observed ( $\ddagger$ ,  $P < 0.005$ ), even for the SNPs that showed moderate associations with rheumatoid arthritis (meta-analysis  $P < 0.4$  in each population).



the two ancestry groups<sup>17</sup>. When the rheumatoid arthritis cases of the Japanese GWAS meta-analysis were stratified into anti-CCP-positive or rheumatoid factor-positive cases ( $n = 3,209$ ) and controls ( $n = 16,891$ ), similar results were observed (data not shown). Nevertheless, most of the SNPs assessed here are not necessarily causal variants, and further fine mapping of the loci is warranted to precisely evaluate the shared genetic predisposition between the populations.

Next, we compared regional associations within each of the loci and identified unique patterns in the *ARID5B* locus at 10q21 (Supplementary Fig. 3). In Japanese, three peaks of association were observed ( $P = 1.0 \times 10^{-8}$  at rs10821944,  $P = 5.7 \times 10^{-8}$  at rs10740069 and  $P = 8.5 \times 10^{-6}$  at rs224311). These three variants were in weak linkage disequilibrium (LD) in Japanese ( $r^2 < 0.10$ ), indicating independent associations with each of the other SNPs that satisfied a region-wide significance threshold of  $P < 3.5 \times 10^{-5}$  (conditional  $P = 4.3 \times 10^{-6}$ ,  $1.7 \times 10^{-5}$  and  $1.8 \times 10^{-5}$ , respectively) (Supplementary Fig. 3). In contrast, there was only one peak of association in Europeans ( $P = 1.2 \times 10^{-6}$  at rs12764378;  $r^2 = 0.59$  with rs10821944 in Europeans), and no additional association was observed in conditional analysis with rs12764378 (the smallest conditional  $P = 2.2 \times 10^{-4}$ ), suggesting that the number of independent associations may be different at this locus in the two populations.

Finally, we conducted polygenic assessment for common variants showing modest associations to rheumatoid arthritis (those not meeting the genome-wide association threshold). This approach has been recognized to be a means to explain a substantial proportion of genetic risk<sup>23</sup>. For the SNPs that were shared between the two meta-analyses but not included in the validated rheumatoid arthritis

susceptibility loci, we adopted LD pruning of the SNPs ( $r^2 < 0.3$ ). We then evaluated the correlation of odds ratios of the SNPs between the two meta-analyses and observed a significant positive correlation ( $r = 0.023$ ,  $P < 1.0 \times 10^{-300}$ ). When the SNPs were stratified according to the  $P$  values in each meta-analysis, significant positive correlations of odds ratios were observed for the SNPs, even for those showing modest association ( $P < 0.4$  in the meta-analysis of Japanese or Europeans;  $r = 0.014$ – $0.36$  for each  $P$  value range,  $P < 0.005$  for each correlation test) (Fig. 3c). Correlations ( $r$ ) of odds ratios observed herein suggest substantial overlap of the genetic risk of rheumatoid arthritis between the two populations, not only in the validated rheumatoid arthritis susceptibility loci but also at the loci showing nonsignificant associations. This suggests the usefulness of a meta-analysis approach involving multiple ancestry groups in identifying additional susceptibility loci.

In summary, we identified multiple new loci associated with rheumatoid arthritis through a large-scale meta-analysis of GWAS in Japanese. Multi-ancestry comparative analysis provided evidence of significant overlap in the genetic risks of rheumatoid arthritis between Japanese and Europeans. Thus, findings from the present study should contribute to the further understanding of the etiology of rheumatoid arthritis.

**URLs.** GARNET consortium, <http://www.twmu.ac.jp/IOR/garnet/home.html>; The BioBank Japan Project (in Japanese), <http://biobankjp.org/>; International HapMap Project, <http://www.hapmap.org/>; PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>; EIGENSTRAT, <http://genepath.med.harvard.edu/~reich/Software.htm>; MACH and mach2dat, <http://www.sph.umich.edu/csg/abecasis/MACH/index>.

html; R statistical software, <http://cran.r-project.org/>; SNAP, <http://www.broadinstitute.org/mpg/snap/index.php>; NCBI GEO database, <http://www.ncbi.nlm.nih.gov/geo/>.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

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

## ACKNOWLEDGMENTS

The authors acknowledge the essential role of the GARNET consortium in developing the study. In this study, the following GARNET members are included: CGM of RIKEN, University of Tokyo, the BioBank Japan Project, Kyoto University and IORRA. We would like to thank all the doctors and staff who participated in sample collection for the RIKEN cohort and the BioBank Japan Project. We thank K. Kobayashi and M. Kitazato for their technical assistance. We thank T. Raj for calculation of composite of multiple signals (CMS). We thank M. Kokubo for DNA extraction, GWAS genotyping and secretarial assistance. We would also like to thank H. Yoshifuji, N. Yukawa, D. Kawabata, T. Nojima, T. Usui and T. Fujii for collecting DNA samples. We thank Y. Katagiri for her technical efforts. We also appreciate the contribution of E. Inoue and other members of the Institute of Rheumatology, Tokyo Women's Medical University, for their efforts on the IORRA cohort. This study was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, the Ministry of Health, Labour and Welfare (MHLW) in Japan, the Japan Society for the Promotion of Science (JSPS), Core Research for Evolutional Science and Technology (CREST), Solution-Oriented Research for Science and Technology (SORST), INSERM and the Okawa Foundation for Information and Telecommunications.

## AUTHOR CONTRIBUTIONS

Y. Okada, C.T., K.I., Y. Kochi and K.O. designed the study and drafted the manuscript. Y. Okada, C.T., K.I., T.K., H.O., N.N., M.T., M.L., K. Tokunaga and M.K. managed genotyping and manipulation of GWAS data. Y. Okada, Y. Kochi, C.T. and K.I. managed genotyping of replication cohorts. Y. Okada, T.K., H.O., E.A.S., A. Takahashi and R.Y. performed statistical analysis. Y. Kochi, A.S., K. Myouzen, T. Sawada, Y. Nishoka, M.Y., T. Matsubara, S.W., R.T. and S.T. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from the BioBank Japan Project and CGM, RIKEN. C.T., K.O., T.K., M.T., K. Takasugi, K.S., A.M., S.H., K. Matsuo, H. Tanaka, K. Tajima and M.L. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from Kyoto University. K.I., T. Suzuki, T.I., Y. Kawamura, H. Tanii, Y. Okazaki and T. Sakaki collected samples and managed phenotype data for the rheumatoid arthritis cohorts from IORRA. Y. Kochi managed the data for the SLE and Graves' disease cohorts. A.S., C.T. and K.I. analyzed the sera of subjects with rheumatoid arthritis. E.A.S., F.A.S.K., P.K.G., J.W., K.A.S., L.P. and R.M.P. managed the data for the rheumatoid arthritis cohorts in European populations. A. Taniguchi, A. Takahashi, K. Tokunaga, M.K., Y. Nakamura, N.K., T. Minori, R.M.P., H.Y., S.M., R.Y., F.M. and K.Y. supervised the overall study.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
2. Kochi, Y. *et al.* A functional variant in *FCRL3*, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
3. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
4. Remmers, E.F. *et al.* *STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
5. Plenge, R.M. *et al.* *TRAF1-C5* as a risk locus for rheumatoid arthritis—a genome-wide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
6. Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
7. Barton, A. *et al.* Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat. Genet.* **40**, 1156–1159 (2008).
8. Suzuki, A. *et al.* Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
9. Gregersen, P.K. *et al.* *REL*, encoding a member of the NF- $\kappa$ B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
10. Kochi, Y. *et al.* A regulatory variant in *CCR6* is associated with rheumatoid arthritis susceptibility. *Nat. Genet.* **42**, 515–519 (2010).
11. Freudenberg, J. *et al.* Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum.* **63**, 884–893 (2011).
12. Terao, C. *et al.* The human *AIRE* gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* **20**, 2680–2685 (2011).
13. Raychaudhuri, S. *et al.* Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**, 1216–1223 (2008).
14. Raychaudhuri, S. *et al.* Genetic variants at *CD28*, *PRDM1* and *CD2/CD58* are associated with rheumatoid arthritis risk. *Nat. Genet.* **41**, 1313–1318 (2009).
15. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
16. Zhemakova, A. *et al.* Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7**, e1002004 (2011).
17. Kurreeman, F. *et al.* Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am. J. Hum. Genet.* **88**, 57–69 (2011).
18. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
19. Yamanaka, H. *et al.* Influence of methotrexate dose on its efficacy and safety in rheumatoid arthritis patients: evidence based on the variety of prescribing approaches among practicing Japanese rheumatologists in a single institute-based large observational cohort (IORRA). *Mod. Rheumatol.* **17**, 98–105 (2007).
20. Yamada, R. *et al.* Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. *Am. J. Hum. Genet.* **68**, 674–685 (2001).
21. Tokuhira, S. *et al.* An intronic SNP in a *RUNX1* binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat. Genet.* **35**, 341–348 (2003).
22. Okada, Y. *et al.* A genome-wide association study identified *AFF1* as a susceptibility locus for systemic lupus erythematosus in Japanese. *PLoS Genet.* **8**, e1002455 (2012).
23. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
24. Shimane, K. *et al.* A single nucleotide polymorphism in the *IRF5* promoter region is associated with susceptibility to rheumatoid arthritis in the Japanese patients. *Ann. Rheum. Dis.* **68**, 377–383 (2009).

<sup>1</sup>Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan. <sup>2</sup>Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. <sup>3</sup>Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan. <sup>4</sup>Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. <sup>5</sup>Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>6</sup>Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan. <sup>7</sup>Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. <sup>8</sup>Broad Institute, Cambridge, Massachusetts, USA. <sup>9</sup>Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands. <sup>10</sup>Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. <sup>11</sup>Department of Rheumatology, Tokyo Medical University Hospital, Tokyo, Japan. <sup>12</sup>Yamanashi Prefectural Central Hospital, Yamanashi, Japan. <sup>13</sup>Department of Orthopaedic Surgery, Yukioka Hospital, Osaka, Japan. <sup>14</sup>Matsubara Mayflower Hospital, Hyogo, Japan. <sup>15</sup>Osaka Minami National Hospital, Osaka, Japan. <sup>16</sup>Department of Orthopedic Surgery, Tottori University, Tottori, Japan. <sup>17</sup>Department of Rheumatology, National Hospital Organization, Sagami Hospital, Kanagawa, Japan. <sup>18</sup>Center for Rheumatic Diseases, Dohgo Spa Hospital, Ehime, Japan. <sup>19</sup>Department of Rheumatology, Niigata Rheumatic Center, Niigata, Japan. <sup>20</sup>Saiseikai Takaoka Hospital, Toyama, Japan. <sup>21</sup>Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. <sup>22</sup>Aichi Cancer Center Hospital and Research Institute, Aichi, Japan. <sup>23</sup>Department of Orthopaedic Surgery, Keio University, Tokyo, Japan. <sup>24</sup>Yokohama Clinic, Warakukai Medical Corporation, Yokohama, Japan. <sup>25</sup>Department of Psychiatry, Mie University School of Medicine, Mie, Japan. <sup>26</sup>Metropolitan Matsuzawa Hospital, Tokyo, Japan. <sup>27</sup>Graduate School of Education, University of Tokyo, Tokyo, Japan. <sup>28</sup>The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA. <sup>29</sup>Rheumatology Unit,

## LETTERS

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



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

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

**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 have 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<sup>+</sup> and CD19<sup>+</sup> 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 Erythematosus 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

**Copyright:** © 2012 Okada et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**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].

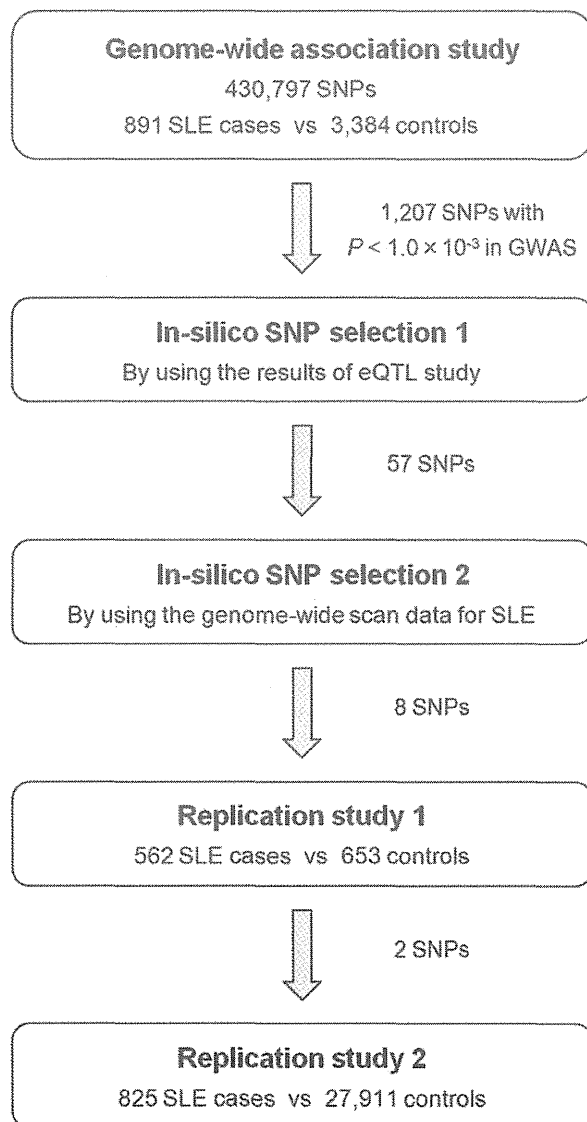
We identified significant associations in six chromosomal loci that satisfied the genome-wide significance threshold of  $P < 5.0 \times 10^{-8}$  (Table 1 and Figure 2A). These loci have been reported to be associated with SLE susceptibility (*STAT4*, *TNFAIP3*, *HIP1*, *BLK*, *ETS1*, and the HLA region)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

### Incorporation of eQTL study into GWAS results

For the selection of SNPs incorporated in the replication studies of the potential association signals, we evaluated cis-eQTL effects of the SNPs using publically available gene expression data [28], and prioritized the results of the GWAS. After applying QC criteria, we evaluated the expression levels of 19,047 probes assayed in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals [29] using Illumina's human whole-genome expression array (WG-6 version 1) [28]. For each of the SNPs included in our GWAS, probes located within  $\pm 300$  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  $\pm 10$  kbp of probes used for the expression analysis, 13.1% were eQTL positive; Table S3).

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





**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  $P < 1.0 \times 10^{-6}$  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;  $P = 4.6 \times 10^{-5}$  and  $P = 0.0094$  in the two individual cohorts, respectively; Table 3, Table S5, and Figure 2B). The combined study for the GWAS ( $P = 1.5 \times 10^{-4}$ ) and the replication studies demonstrated significant associations of rs340630 that satisfied the genome-wide significance threshold ( $P = 8.3 \times 10^{-9}$ , 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<sup>+</sup> and CD19<sup>+</sup> 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<sup>+</sup> and CD19<sup>+</sup> 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 <sup>a</sup> | Chr | Position (bp) | Cytoband | Gene           | Allele <sup>b</sup> | No. subjects |         | Allele 1 freq. |         | OR (95%CI)       | P                     |
|-------------------|-----|---------------|----------|----------------|---------------------|--------------|---------|----------------|---------|------------------|-----------------------|
|                   |     |               |          |                |                     | Case         | Control | Case           | Control |                  |                       |
| rs10168266        | 2   | 191,644,049   | 2q32     | <i>STAT4</i>   | T/C                 | 891          | 3,384   | 0.37           | 0.27    | 1.59 (1.42–1.78) | $2.7 \times 10^{-16}$ |
| rs9501626         | 6   | 32,508,322    | 6p21     | HLA region     | A/C                 | 891          | 3,381   | 0.20           | 0.12    | 1.86 (1.62–2.13) | $1.0 \times 10^{-18}$ |
| rs2230926         | 6   | 138,237,759   | 6q23     | <i>TNFAIP3</i> | G/T                 | 891          | 3,377   | 0.11           | 0.069   | 1.75 (1.47–2.08) | $1.9 \times 10^{-10}$ |
| rs6964720         | 7   | 75,018,280    | 7q11     | <i>HIP1</i>    | G/A                 | 891          | 3,384   | 0.25           | 0.19    | 1.43 (1.27–1.63) | $1.3 \times 10^{-8}$  |
| rs2254546         | 8   | 11,381,089    | 8p23     | <i>BLK</i>     | G/A                 | 891          | 3,384   | 0.78           | 0.72    | 1.42 (1.61–1.25) | $4.1 \times 10^{-8}$  |
| rs6590330         | 11  | 127,816,269   | 11q24    | <i>ETS1</i>    | A/G                 | 891          | 3,368   | 0.48           | 0.39    | 1.44 (1.30–1.60) | $1.3 \times 10^{-11}$ |

<sup>a</sup>SNPs that satisfied the threshold of  $P < 5.0 \times 10^{-8}$  were indicated.

<sup>b</sup>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<sup>+</sup> and CD19<sup>+</sup> 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

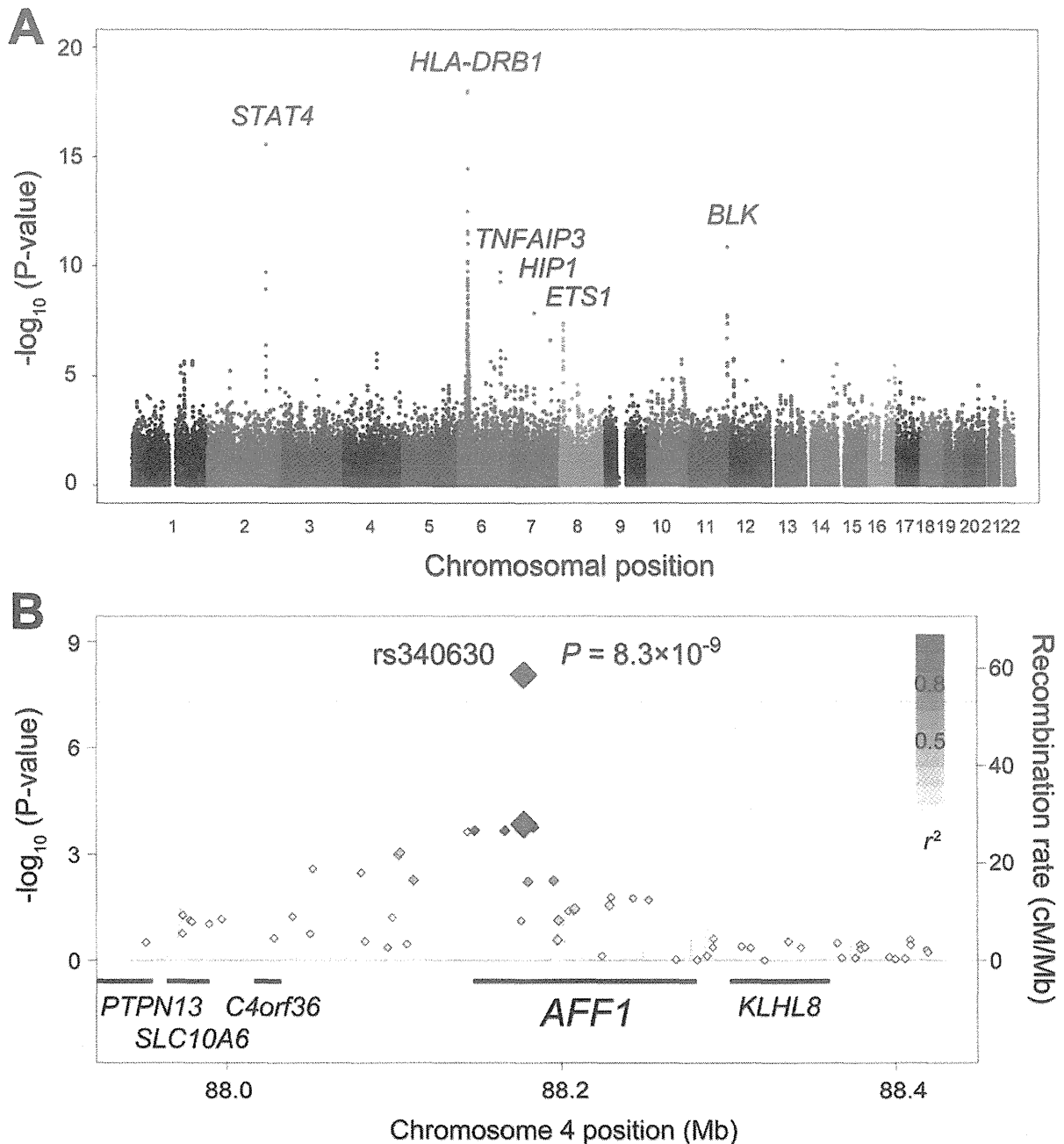
between the studies would be difficult due to these issues, we adopted the complementary approach that referred the association signals of the multiple SNPs in the pooled analysis for each of the single SNPs in the GWAS, taking account of LD and physical distances between the SNPs. However, there would 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 <sup>a</sup><br>1/2 | Allele 1 freq. |         | OR (95%CI)       | P                     | eQTL <sup>b</sup> | Identified by the studies in <sup>c</sup> |        |
|------------|-----|---------------|----------|--------------------|----------------------------|----------------|---------|------------------|-----------------------|-------------------|---|--------|
|            |     |               |          |                    |                            | Case           | Control |                  |                       |                   | Caucasians                                | Asians |
| rs2205960  | 1   | 171,458,098   | 1q25     | <i>TNFSF4</i>      | T/G                        | 0.23           | 0.18    | 1.35 (1.19–1.54) | $3.0 \times 10^{-6}$  | +                 |   |        |
| rs3024505  | 1   | 205,006,527   | 1q32     | <i>IL10</i>        | A/G                        | 0.019          | 0.014   | 1.34 (0.90–2.00) | 0.15                  |                   |   |        |
| rs13385731 | 2   | 33,555,394    | 2p22     | <i>RASGRP3</i>     | C/T                        | 0.90           | 0.87    | 1.37 (1.15–1.64) | $6.0 \times 10^{-4}$  | +                 |   | +      |
| rs10168266 | 2   | 191,644,049   | 2q32     | <i>STAT4</i>       | T/C                        | 0.37           | 0.27    | 1.59 (1.42–1.78) | $2.7 \times 10^{-16}$ |                   | +   |        |
| rs6445975  | 3   | 58,345,217    | 3p14     | <i>PXK</i>         | G/T                        | 0.25           | 0.23    | 1.09 (0.96–1.23) | 0.18                  | +                 | +   |        |
| rs10516487 | 4   | 102,970,099   | 4q24     | <i>BANK1</i>       | G/A                        | 0.91           | 0.89    | 1.28 (1.07–1.53) | 0.0070                |                   | +   |        |
| rs10036748 | 5   | 150,438,339   | 5q33     | <i>TNIP1</i>       | T/C                        | 0.75           | 0.72    | 1.16 (1.03–1.31) | 0.014                 |                   |   | +      |
| rs9501626  | 6   | 32,508,322    | 6p21     | <i>HLA-DRB1</i>    | A/C                        | 0.20           | 0.12    | 1.86 (1.62–2.13) | $1.0 \times 10^{-18}$ |                   | +   |        |
| rs548234   | 6   | 106,674,727   | 6q21     | <i>PRDM1</i>       | C/T                        | 0.40           | 0.34    | 1.30 (1.16–1.44) | $2.3 \times 10^{-6}$  | +                 | +   |        |
| rs2230926  | 6   | 138,237,759   | 6q23     | <i>TNFAIP3</i>     | G/T                        | 0.11           | 0.069   | 1.75 (1.47–2.08) | $1.9 \times 10^{-10}$ | +                 | +   |        |
| rs849142   | 7   | 28,152,416    | 7p15     | <i>JAZF1</i>       | C/T                        | 0.999          | 0.999   | 2.72 (0.25–29.8) | 0.41                  |                   | +   |        |
| rs4917014  | 7   | 50,276,409    | 7p12     | <i>IKZF1</i>       | T/G                        | 0.58           | 0.53    | 1.24 (1.11–1.38) | $8.1 \times 10^{-5}$  |                   |   | +      |
| rs6964720  | 7   | 75,018,280    | 7q11     | <i>HIP1</i>        | G/A                        | 0.25           | 0.19    | 1.43 (1.27–1.62) | $1.3 \times 10^{-8}$  |                   |   | +      |
| rs4728142  | 7   | 128,361,203   | 7q32     | <i>IRF5</i>        | A/G                        | 0.16           | 0.11    | 1.48 (1.28–1.72) | $2.4 \times 10^{-7}$  | +                 | +   |        |
| rs2254546  | 8   | 11,381,089    | 8p23     | <i>BLK</i>         | G/A                        | 0.78           | 0.72    | 1.42 (1.25–1.61) | $4.1 \times 10^{-8}$  | +                 | +   |        |
| rs1913517  | 10  | 49,789,060    | 10q11    | <i>WDFY4</i>       | A/G                        | 0.32           | 0.28    | 1.20 (1.07–1.35) | 0.0013                |                   |   | +      |
| rs4963128  | 11  | 579,564       | 11p15    | <i>KIAA1542</i>    | T/C                        | 0.98           | 0.97    | 1.58 (1.03–2.44) | 0.038                 | +                 | +   |        |
| rs2732552  | 11  | 35,041,168    | 11p13    | <i>PDHX, CD44</i>  | 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) | $7.3 \times 10^{-4}$  |                   |   | +      |
| rs6590330  | 11  | 127,816,269   | 11q24    | <i>ETS1</i>        | A/G                        | 0.48           | 0.39    | 1.44 (1.30–1.60) | $1.3 \times 10^{-11}$ |                   |   | +      |
| rs1385374  | 12  | 127,866,647   | 12q24    | <i>SLC15A4</i>     | T/C                        | 0.19           | 0.16    | 1.21 (1.06–1.38) | 0.0057                |                   |   | +      |
| rs7329174  | 13  | 40,456,110    | 13q14    | <i>ELF1</i>        | G/A                        | 0.30           | 0.25    | 1.32 (1.18–1.49) | $2.2 \times 10^{-6}$  |                   |   | +      |
| 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    | <i>ITGAM</i>       | 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    | <i>HIC2,UBE2L3</i> | T/C                        | 0.52           | 0.48    | 1.20 (1.08–1.33) | $6.1 \times 10^{-4}$  |                   | +   |        |

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

### Genotyping and quality control

In GWAS, 946 SLE cases and 3,477 controls were genotyped using Illumina HumanHap610-Quad and Illumina Human-

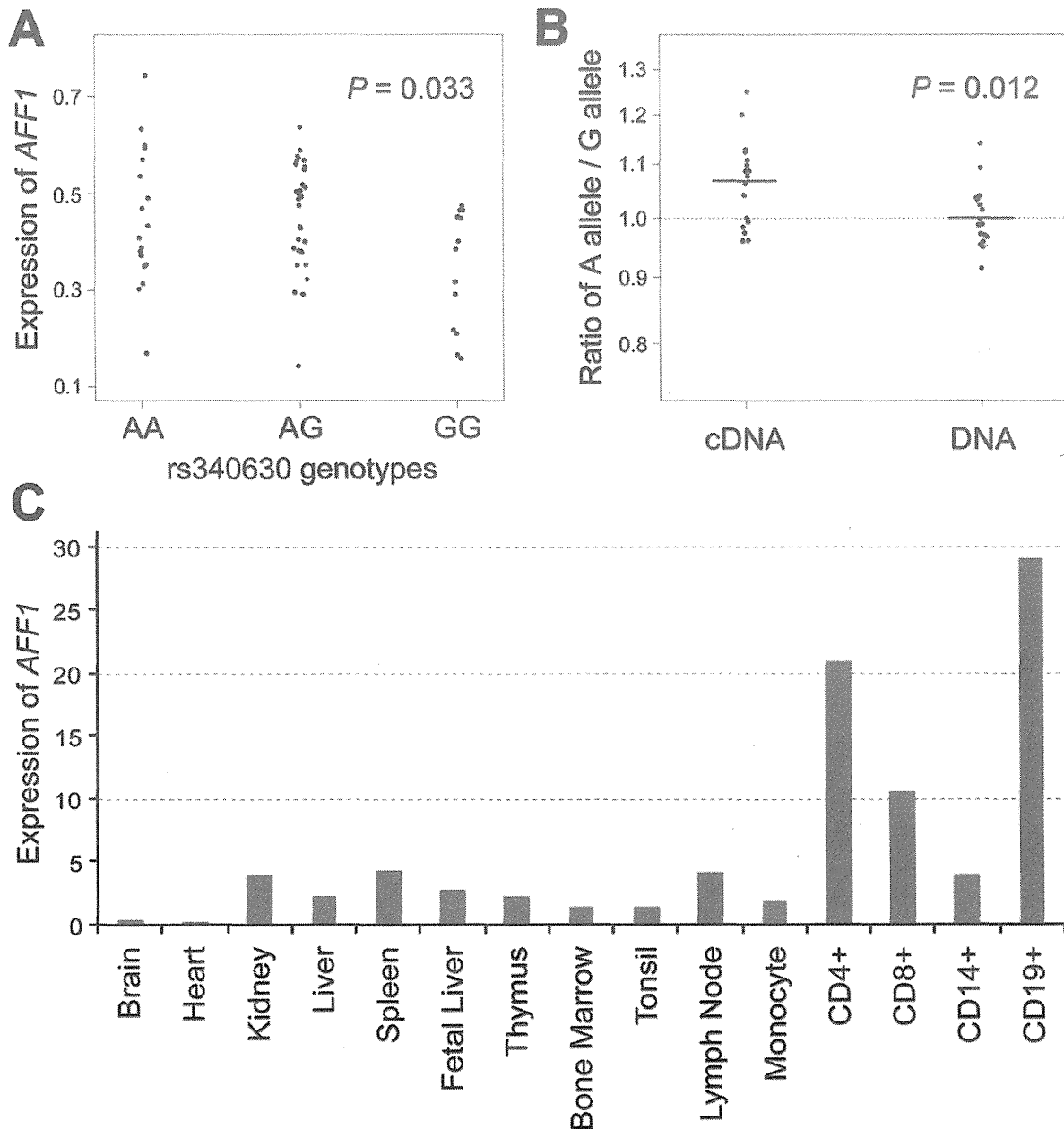
Hap550v3 Genotyping BeadChips (Illumina, CA, USA), respectively. After the exclusion of 47 SLE cases and 92 controls with call rates <0.98, SNPs with call rates <0.99 in SLE cases or controls,

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

| rsID     | Chr | Position (bp) | Cytoband | Gene        | Allele<br>1/2 | Stage               | No. subjects |         | Allele 1 freq. |         | OR (95%CI)       | P                    | eQTL <sup>a</sup> |
|----------|-----|---------------|----------|-------------|---------------|---------------------|--------------|---------|----------------|---------|------------------|----------------------|-------------------|
|          |     |               |          |             |               |                     | Case         | Control | Case           | Control |                  |                      |                   |
| rs340630 | 4   | 88,177,419    | 4q21     | <i>AFF1</i> | A/G           | GWAS                | 891          | 3,383   | 0.56           | 0.51    | 1.22 (1.10–1.36) | $1.5 \times 10^{-4}$ | +                 |
|          |     |               |          |             |               | Replication study 1 | 550          | 646     | 0.57           | 0.49    | 1.40 (1.19–1.64) | $4.6 \times 10^{-5}$ |                   |
|          |     |               |          |             |               | 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) | $8.3 \times 10^{-9}$ |                   |

<sup>a</sup>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  $< 1.0 \times 10^{-6}$  in controls, or SNPs with ambiguous cluster plots were excluded. Finally, 430,797 SNPs for 891 SLE cases and 3,384 controls were obtained. Genotyping of SNPs in replication studies was performed using TaqMan Assay or Illumina HumanHap610-Quad Genotyping BeadChip (Illumina, CA, USA).

### Association analysis of the SNPs

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

### 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  $\pm 300$  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  $P < 1.0 \times 10^{-4}$  in GWAS, or the SNPs that satisfied  $1.0 \times 10^{-4} \leq P < 1.0 \times 10^{-3}$  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  $\pm 100$  kbp of SNPs showing association signals in the pooled analysis ( $P < 0.01$ ), it would be selected. SNPs that satisfied  $P < 1.0 \times 10^{-6}$  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  $\mu$ g) 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-GACCCGCGCGCCCAAC-TAMRA-3' and 5'-FAM-CGAA-GACCCGCGCGCCCAA-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  $1.0 \times 10^{-15}$  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)

## References

- Lipsky PE (2001) Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat Immunol* 2: 764–766.
- 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.
- Sigurðsson 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.
- 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.
- 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.
- 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.
- Cunningham 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.
- 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.
- 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, PTK, KIAA1542 and other loci. *Nat Genet* 40: 204–210.
- 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.
- 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.
- Graham RR, Cosapas 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.
- Musone SL, Taylor KE, Lu TT, Nithilam J, Ferreira RC, et al. (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat Genet* 40: 1062–1064.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Danchenko N, Satia JA, Anthony MS (2006) Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* 15: 308–318.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. *Nat Genet* 39: 1217–1224.
- The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789–796.



30. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, McQuist S, et al. (2007) Identification of the genetic basis for complex disorders by use of pooling-based genome-wide 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, Shores 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.

