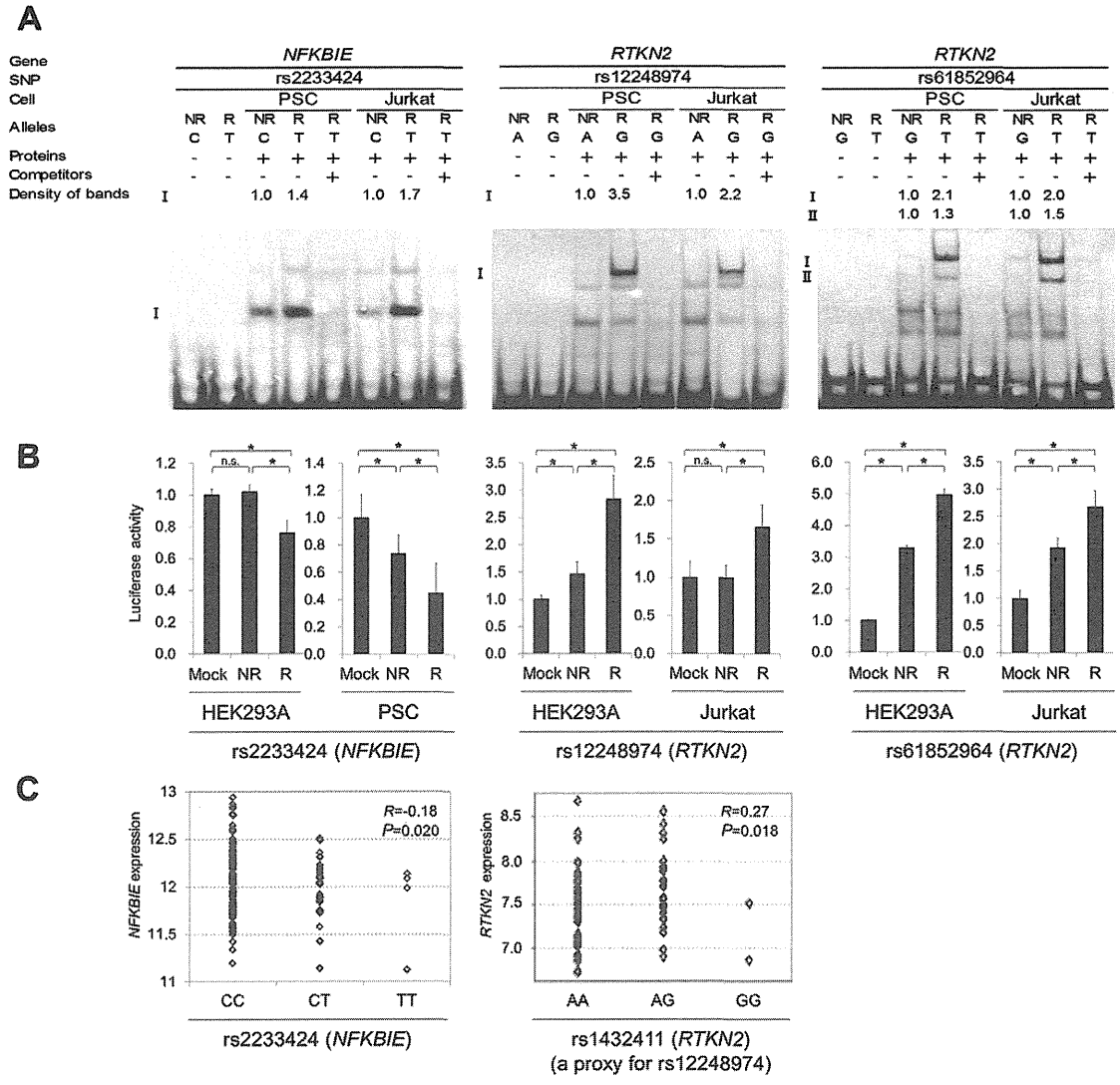


Figure 5. Evaluation of candidate regulatory SNPs *in vitro*. (A) Binding of nuclear factors from lymphoblastoid B-cells (PSC cells) and Jurkat cells to the 31-bp sequences around each SNP was evaluated by EMSA. Unlabeled probes in 200-fold excess as compared to the labeled probes were used for the competition experiment. The densities of the bands were quantified and normalized to that of the risk allele. rs2233424 in *NFKBIE* (C(NR)/T(R)) (left), rs12248974 (A(NR)/G(R)) (middle) and rs61852964 (G(NR)/T(R)) (right) in *RTKN2*. (B) Transcriptional activities were evaluated by luciferase assays. Each 31-bp oligonucleotide was inserted into the pGL4.24[*Luc2P*/minP] vector. Luc, luciferase; minP, minimal promoter. Transcfection was performed with HEK293A (for all the SNPs), PSC cells (for rs2233424), and Jurkat cells (for rs12248974 and rs61852964). rs2233424 (left), rs12248974 (middle), and rs61852964 (right). Data represent the mean ± s.d. Each experiment was performed in sextuplicate and independently repeated three times. *$P<0.05$ by Student's *t*-test. n.s.: not significant. (C) Liner regression analysis of the relationship between SNP genotype and gene expression level. *NFKBIE* expression data in lymphoblastoid B-cell lines of HapMap individuals (JPT+CHB, CEU and YRI; $n=151$), and *RTKN2* expression data in primary T cells from umbilical cords of Western European individuals ($n=85$) were used. The x-axis shows the SNP genotypes and the y-axis represents the $\log_2$-transformed gene expression level. R: correlation coefficient between SNP genotype and gene expression. Rs2233424 genotypes and *NFKBIE* expression level (left). The genotype classification by population: JPT+CHB, CC = 52, CT = 1; CEU, CC = 35, CT = 2; YRI, CC = 32, CT = 2, TT = 4. Rs1432411 genotypes and *RTKN2* expression level (right). Rs1432411 was used as a proxy SNP of rs12248974 ($r^2=0.97$).
doi:10.1371/journal.pgen.1002949.g005

reticuloendotheliosis viral oncogene homolog (*REL* [MIM 164910]) [5], TNF receptor-associated factor 1 (*TRAF1* [MIM 601711]) [3], and CD40 molecule TNF receptor superfamily member 5 (*CD40* [MIM 109535]) [38].

In conclusion, we identified *NFKBIE* and *RTKN2* as genetic risk factors for RA. Considering the allelic effect of both genes, enhanced NF-κB activity may play a role in the pathogenesis of the disease. Because NF-κB regulates the expression of numerous genes, including inflammatory and immune response mediators, NF-κB and its regulators identified by GWAS are promising targets for the treatment of RA.

Materials and Methods

Ethics statement

All subjects were of Japanese origin and provided written informed consent for participation in the study, which was approved by the ethical committees of the institutional review boards.

Subjects

A total of 7,907 RA cases, 657 SLE cases, 1,783 GD cases, and 35,362 control subjects were enrolled in the study through medical

institutes in Japan under the support of the BioBank Japan Project, Center for Genomic Medicine at RIKEN, the University of Tokyo, Tokyo Women's Medical University, and Kyoto University. The same case and control samples were used in the previous meta-analysis of GWASs in the Japanese population (Table S1) [15] . RA and SLE subjects met the revised American College of Rheumatology (ACR) criteria for RA [39]. Diagnosis of individuals with GD was established on the basis of clinical findings and results of the routine examinations for circulating thyroid hormone and thyroid-stimulating hormone concentrations, thyroid-stimulating hormone receptors, ultrasonography, $^{[99m]}TCO_4^-$ (or [^{123}I]) uptake, and thyroid scintigraphy. DNAs were extracted from peripheral blood cells using a standard protocol. Total RNAs were also extracted from PBMCs of healthy individuals ($n = 20$) using an RNeasy kit (QIAGEN, Valencia, CA, USA). Details of the samples are summarized in Table S1.

Genotyping and quality control

In the GWAS, RA cases and controls were genotyped using Illumina Human610-Quad and Illumina Human 550v3 Genotyping BeadsChips (Illumina, San Diego, CA, USA), respectively, and quality control of genotyping was performed as described previously [6]. For replication study of candidate loci, a landmark SNP was selected from each locus that satisfied $5\times10^{-8}<P_{GWAS}<5\times10^{-5}$ in the GWAS. If multiple candidate SNPs existed within ±100 kb, the SNP with the lowest *P*-value was selected. All case subjects in the replication study and both case and control subjects in the validation study of candidate causal variants were genotyped using TaqMan SNP genotyping assays (Table S12) (Applied Biosystems, Foster City, CA, USA) with an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Because of the availability of DNA samples, only a part of the control subjects were genotyped for the validation study ($n = 3{,}290$, 97.3%). To enlarge the number of subjects and enhance statistical power for replication studies, we used genotype data obtained from other GWAS projects genotyped using the Illumina platforms for the replication control panels (Table S1). All SNPs were successfully genotyped with call rates >0.98 and were in Hardy-Weinberg equilibrium (HWE) in control subjects ($P>0.05$ as examined by χ^2 test), except for rs2233434, which displayed a deviation from HWE ($P = 0.00091$). To evaluate possible genotyping biases between the platforms, we also genotyped rs2233434 and rs3125734 by TaqMan assays for randomly selected subjects genotyped using other genotyping platforms ($n = 376$), yielding high concordance rates of ≥0.99.

Association analysis

The associations of the SNPs were tested with the Cochran-Armitage trend test. Combined analysis was performed with the Mantel-Haenszel method. Haplotype association analysis and haplotype-based conditional association analysis were performed using Haploview v4.2 and the PLINK v1.07 program (see URLs) [40], respectively. The SNPs that were not genotyped in the GWAS were imputed using MACH 1.0.16 (see URLs), with genotype data from the 1000 Genome Project (JPT, CHB, and Han Chinese South (CHS): 177 individuals) as references (August 2010 release) [41]. All the imputed SNPs demonstrated *Rsq* values more than 0.60.

DNA re-sequencing

Unknown variants in the coding sequences of *NFKBIE* and *RTKN2* were revealed by directly sequencing the DNA of 48 individuals affected with RA. DNA fragments were amplified with the appropriate primers (Table S13). Purification of PCR products was performed with Exonuclease I (New England Biolabs, Ipswich, MA, USA) and shrimp alkaline phosphatase (Promega, Madison, WI, USA). The amplified DNAs were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), and signals were detected using an ABI 3700 DNA Analyzer (Applied Biosystems).

Construction of haplotype-specific expression vectors

The full coding regions were amplified using cDNAs prepared from an Epstein-Barr virus-transfected lymphoblastoid B-cell line (Pharma SNP Consortium (PSC), Osaka, Japan) for *NFKBIE* (NM_004556.2) and from Jurkat cells (American Type Culture Collection (ATCC), Rockville, MD, USA) for *RTKN2* (NM_145307.2) with appropriate primers (Table S14) and DNA polymerases. PCR products were inserted into the pcDNA3.1D/V5-His-TOPO vector (Invitrogen, Camarillo, CA, USA) using the TaKaRa Ligation kit ver. 2.1 (Takara Bio Inc, Shiga, Japan), and mutagenized using the AMAP Multi Site-Directed Mutagenesis Kit (MBL, Nagoya, Japan). Each construct was then transformed into Jet Competent *Escherichia coli* cells (DH5α) (BioDynamics Laboratory Inc., Tokyo, Japan). These plasmids were purified using an Endofree Plasmid Maxi Kit (QIAGEN) after confirmation of the sequence.

NF-κB reporter assay

Human embryonic kidney (HEK) 293A cells (Invitrogen) were cultured in Dulbecco's modified Eagle's medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% fetal bovine serum (BioWest, Nuaillé, France), 1% penicillin/streptomycin (Invitrogen), and 0.1 mM MEM Non-Essential Amino Acids (Invitrogen). Various doses of the haplotype-specific expression vector (0.0025–0.02 μg for *NFKBIE* and 0.1–0.8 μg for *RTKN2*), pGL4.32[*luc2P*/NF-κB-RE/Hygro] vector (Promega) (0.05 μg and 0.0125 μg, respectively), and pRL-TK vector (an internal control for transfection efficiency) (0.45 μg and 0.15 μg, respectively) were transfected into the HEK293A cells using the Lipofectamine LTX transfection reagent (Invitrogen) according to the manufacturer's protocol. The total amounts of DNAs were adjusted with empty pcDNA3.1 vector. After 22 h, cells were incubated with 1 ng/ml TNF-α (Sigma) for 2 h or with medium alone. Cells were collected, and luciferase activity was measured using a Dual-Luciferase Reporter Assay system (Promega) and a GloMax-Multi+ Detection System (Promega). Each experiment was independently repeated three times, and sextuplicate samples were assayed each time.

Western blotting

After 24 h of transfection as described for the NF-κB reporter assay, cells were lysed in NP-40 lysis buffer (150 mM NaCl, 1% NP-40, 50 mM Tris-HCl at pH 8.0, and a protease inhibitor cocktail), and incubated on ice for 30 min. After centrifugation, the supernatant fraction was collected and 4×Sodium dodecyl sulfate (SDS) sample buffer was added. After denaturation at 95°C for 5 min, proteins were analyzed by SDS-polyachrylamide gel electrophoresis (PAGE) on a 5% to 20% gradient gel (Wako, Osaka, Japan) and were transferred to polyvinylidene difluouride (PVDF) membranes (Millipore, Billerica, MA, USA). Target proteins on the membrane were probed with antibodies (mouse anti-V5 tag (Invitrogen), anti-β-actin-HRP (an internal control), and goat anti-mouse IgG2a-HRP (Santa Cruz Biotechnology, Santa Cruz, CA, USA)), visualized using enhanced chemiluminescence (ECL) detection reagent (GE Healthcare, Pollards Wood, UK), and detected using a LAS-3000 mini lumino-image analyzer

(Fujifilm, Tokyo, Japan). Band intensities were measured using MultiGauge software (Fujifilm).

Allele-specific transcript quantification (ASTQ) analysis

ASTQ analysis was performed as previously described [42]. Total RNAs and genomic DNAs were extracted from PBMCs and lymphoblastoid B-cell lines. cDNAs were synthesized using TaqMan reverse transcription reagents (Applied Biosystems). We selected SNPs (rs2233434 (A/G) for *NFKBIE* and rs3125734 (C/T) for *RTKN2*) as target SNPs. Allele-specific gene expression was measured by TaqMan SNP genotyping probes for these SNPs (Applied Biosystems). To make a standard curve, we selected two individuals that had homozygous genotypes of each target SNP. We mixed these DNAs at nine different ratios and detected the intensities. The $\log_2$ of (risk allele/non-risk allele intensity) for each SNP was plotted against the $\log_2$ of mixing homozygous DNAs. We generated a standard curve (linear regression line; $y = ax+b$), where y is the $\log_2$ of (risk allele/non-risk allele intensity) at a given mixing ratio, x is the $\log_2$ of the mixing ratio, a is the slope, and b is the intercept. We then measured the allelic ratio for each cDNA and genomic DNA from each individual by real-time TaqMan PCR. Based on a standard curve, we calculated the allelic ratio of cDNAs and genomic DNAs. Intensities were detected using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems).

Electrophoretic mobility shift assays (EMSA)

EMSA and preparation of nuclear extract from lymphoblastoid B-cell lines and Jurkat cells were performed as previously described [43]. Cells were cultured in RPMI-1640 medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. Following stimulation with 50 ng/ml phorbol myristate acetate (Sigma-Aldrich) for 2 h, cells were collected and suspended in buffer A (20 mM HEPES at pH 7.6, 20% glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA at pH 8.0, 1 mM DTT, 0.1% NP-40, and a protease inhibitor cocktail) for 10 min on ice. After centrifugation, the pellets were resuspended in buffer B (which contains buffer A with 500 mM NaCl). Following incubation on ice for 30 min and centrifugation to remove cellular debris, the supernatant fraction containing nuclear proteins was collected. Oligonucleotides (31-bp) were designed that corresponded to genomic sequences surrounding the SNPs (Table S15). Single-stranded oligonucleotide probes were labeled using a Biotin 3′ End DNA Labeling Kit (Pierce Biotechnology, Rockford, IL, USA), and sense and antisense oligonucleotides were then annealed. DNA-protein interactions were detected using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology). The DNA-protein complexes were separated on a non-denaturing 5% polyachrylamide gel in 1×TBE (Tris-borate-EDTA) running buffer for 60 min at 150 V. The DNA-protein complexes were then transferred from the gel onto a nitrocellulose membrane (Ambion, Carlsbad, CA, USA), and were cross-linked to the membrane by exposure to UV light. Signals were detected using a LAS-3000 mini lumino-image analyzer (Fujifilm). Allelic differences were analyzed using MultiGauge software (Fujifilm) by measuring the intensity of the bands.

Luciferase assay

Oligonucleotides (31-bp) were designed as described for the EMSAs (Table S15), and complementary sense and antisense oligonucleotides were annealed. To construct luciferase reporter plasmids, pGL4.24[*luc2P*/minP] vector (Promega) was digested with restriction enzymes (XhoI and BglII) (Takara Bio Inc), and annealed oligonucleotide was ligated into a pGL4.24 vector upstream of the minimal promoter. HEK293A ($n = 2.5\times10^5$), lymphoblastoid B-cell lines ($n = 2.0\times10^6$) and Jurkat ($n = 5.0\times10^5$) cells were transfected with the allele-specific constructs (0.4 μg, 1.8 μg and 2.5 μg, respectively) and the pRL-TK vector (0.1 μg, 0.2 μg and 0.25 μg, respectively) using the Lipofectamine LTX transfection reagent (for HEK293A and Jurkat cells) and Amaxa nucleofector kit (Lonza, Basel, Switzerland) (for lymphoblastoid B-cell lines). Cells were collected, and luciferase activity was measured as described for the NF-κB reporter assay. Each experiment was independently repeated three times and sextuplicate samples were assayed each time.

Correlation analysis between gene expression and genotypes

The expression data in lymphoblastoid B-cell lines derived from HapMap individuals ($n = 210$; JPT, CHB, CEU, and YRI) and in primary T cells from umbilical cords of Western European individuals ($n = 85$) from the database of the Gene Expression Variation (Genevar) project were used. SNP genotypes were obtained from HapMap and 1000 Genome Project databases. The expression levels were regressed with the genotype in a liner model. The statistical significance of regression coefficients was tested using Student's *t*-test.

Statistical analysis

We used χ^2 contingency table tests to evaluate the significance of differences in allele frequency in the case-control subjects. We defined haplotype blocks using the solid spine of LD definition of Haploview v4.2, and estimated haplotype frequency and calculated pairwise LD indices (r^2) between pairs of polymorphisms using the Haploview program. Luciferase assay data and ASTQ analysis data were analyzed by Student's *t*-test.

Web resources

The URLs for data presented herein are as follows:

PLINK, http://pngu.mgh.harvard.edu/~purcekk/plink

MACH, http://www.sph.umich.edu/csg/abecasis/mach/

UCSC Genome Browser, http://genome.ucsc.edu/;

Genevar, http://www.sanger.ac.uk/resources/software/genevar/

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

1000 Genome Project, http://www.1000genomes.org

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

Supporting Information

Figure S1 NF-kB activity was influenced by nsSNPs in *NFKBIE*. NF-κB activities were evaluated by luciferase assays. Allele specific construct, pGL4.32[*luc2P*/NF-κB-RE] luciferase vector, and pRL-TK vector were transfected into HEK293A cells. Four haplotypes (rs2233434-rs2233433; A-C, G-C, A-T, and G-T) were examined. (rs2233434: A = non-risk (NR), G = risk (R); rs2233433: C = NR, T = R). Twenty-two hours after transfection, cells were stimulated with medium alone (A) or TNF-α (B) for 2 h. Data represent the mean ± s.d. Each experiment was performed in sextuplicate, and experiments were independently repeated three times. $*P<0.05$ and $**P<1.0\times10^{-5}$ by Student's *t*-test. n.s.: not significant.
(TIF)

Figure S2 Allelic imbalance of expression in *NFKBIE*. ASTQ was performed using samples from individuals heterozygous for rs2233434 (G/A) in *NFKBIE*. Genomic DNAs and cDNAs were extracted from lymphoblastoid B cells ($n = 9$). The y-axis shows the $\log_2$ ratio of the transcript amounts in target SNPs (risk allele/non-risk

allele). The top bar of the box-plot represents the maximum value and the lower bar represents the minimum value. The top of box is the third quartile, the bottom of box is the first quartile, and the middle bar is the median value. The circle is an outlier. *$P = 5.3 \times 10^{-4}$ by Student's *t*-test.
(TIF)

Figure S3 SNP selection using *in silico* analysis in the *NFKBIE* region. Step 1: Definition of the target region. *P*-values of the SNPs in the GWAS (top) and genomic structure (middle), and the *D'*-based LD map (bottom). The green diamond shapes represent the -$\log_{10}$ of the Cochran-Armitage trend *P*-values. The dashed line indicates the significance threshold ($P < 1 \times 10^{-3}$). The LD map was drawn based on genotype data of the 1000 Genome Project (JPT, CHB and CHS: 177 samples) using Haploview software v4.2. LD blocks were defined by the solid spine method. The red box (top) represents the target region of the *in silico* analysis (Chr6: 44,336,140-44,394,125). Step 2: Target SNPs were extracted from public databases (HapMap and 1000 Genome Project). SNPs with MAF >0.05 were selected. Step 3: Evaluation of regulatory potential. Step 3a: The regulatory potential (RP) score was calculated for sequences surrounding the SNPs by ESPERR (evolutionary and sequence pattern extraction through reduced representations) method. SNPs with RP score >0.1 were selected. Step 3b: Subsequently, SNPs within the predicted, regulatory genomic elements were selected by using ChIP-seq data of transcription factor binding sites (Txn factor), histone modification sites (CTCF binding, H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9ac) or DNase-seq data of DNase I hypersensitivity sites (DNase HS). ChIP-seq data and DNase-seq data used the signals derived from GM12878 EBV-transformed B cells. All these analyses of Steps 2 to 3 were performed by using the UCSC genome browser. Step 4: Evaluation of disease association. Association data of both genotyped (green diamonds) and imputed (black diamonds) SNPs in the GWAS samples were used. Red triangles represent 14 extracted SNPs *in silico*. The dashed line indicates the significance threshold ($P < 0.05$).
(TIF)

Figure S4 SNP selection using *in silico* analysis in the *RTKN2* region. SNP selection in the *RTKN2* region was performed the same as in the case of the *NFKBIE* region as described in Figure S3, except that we used DNase-seq data derived from Th1, Th2, and Jurkat cells in addition to GM12878 EBV-transformed B cells.
(TIF)

Figure S5 Results of EMSAs for candidate regulatory SNPs. Binding affinities of nuclear factors from lymphoblastoid B-cells (PSC cells) and Jurkat cells to the 31-bp sequences around each allele of the candidate regulatory SNPs were evaluated by EMSA. Nuclear factors from PSC cells were used for *NFKBIE*, and Jurkat cells were used for *RTKN2*. 14 SNPs in *NFKBIE* (A) and 10 SNPs in *RTKN2* (B) were tested. NR: non-risk allele; R: risk allele. Arrows indicate bands showing allelic differences in each SNP.
(TIF)

Figure S6 Luciferase assays for regulatory SNPs. Transcriptional activities of the 31-bp genomic sequences around the SNPs were evaluated by luciferase assays. Each oligonucleotide was inserted into the pGL4.24[*luc2P*/minP] vector upstream of the minimal promoter (minP), and allele-specific constructs were transfected into HEK293A cells. Relative luciferase activity is expressed as the ratio of luciferase activity of each allele-specific construct to the luciferase activity of the mock construct. Data represent the mean ± s.d. Each experiment was independently repeated three times, and each sample was measured in sextuplicate. *$P < 1 \times 10^{-3}$ by Student's *t*-test. n.s.: not significant. (A) rs2233434 and rs77986492 in the *NFKBIE* region. (B) rs3864793, rs1864836, rs4979765, and rs4979766 in the *RTKN2* region. NR: non-risk allele; R: risk allele.
(TIF)

Figure S7 The correlation between *NFKBIE* expression and rs2233434 and rs77986492 genotypes. Linear regression analysis of the relationship between SNP genotypes and *NFKBIE* expression. Gene expression data from EBV-transformed lymphoblastoid B cell lines of HapMap individuals (JPT+CHB, CEU, and YRI). (A) rs2233434 ($n = 204$) and (B) rs77986492 ($n = 152$). The genotype classification by population: rs2233434 (JPT+CHB, AA = 61, AG = 28, GG = 1; CEU, AA = 52, AG = 2; YRI, AA = 53, AG = 72) and rs77986492 (JPT+CHB, CC = 52, CT = 24; CEU, CC = 35, CT = 2; YRI, CC = 38, CT = 1). The x-axis shows SNP genotypes and the y-axis represents the $\log_2$-transformed *NFKBIE* expression level. *R*: the correlation coefficient between *NFKBIE* expression and SNP genotype.
(TIF)

Figure S8 The correlation between *RTKN2* expression and rs3852694 genotypes. Linear regression analysis of the relationship between the rs3852694 genotype and *RTKN2* expression. Rs3852694 was used as a proxy SNP of rs1864836 ($r^2 = 1.0$). Gene expression data in primary T cells from umbilical cords of Western European individuals ($n = 85$) were presented by using Genevar software. The x-axis shows the rs3852694 genotypes (AA, AG, GG) and the y-axis represents the $\log_2$-transformed *RTKN2* expression level. *R*: the correlation coefficient between *RTKN2* expression and rs3852694 genotype.
(TIF)

Table S1 Summary of samples.
(DOC)

Table S2 Association results of the GWAS and 1st replication study.
(DOC)

Table S3 Association analysis of *NFKBIE* and *RTKN2* with autoimmune diseases.
(DOC)

Table S4 Association analysis of nsSNPs with RA.
(DOC)

Table S5 Haplotype association study of nsSNPs in *NFKBIE*.
(DOC)

Table S6 Haplotype association study of nsSNPs in *RTKN2*.
(DOC)

Table S7 Predicting the effects of nsSNPs on protein function.
(DOC)

Table S8 Association analysis of candidate rSNPs with RA.
(DOC)

Table S9 Haplotype association study of candidate causal SNPs in *NFKBIE*.
(DOC)

Table S10 Haplotype association study of candidate causal SNPs in *RTKN2*.
(DOC)

Table S11 The conditional haplotype-based association analysis of candidate causal SNPs in *RTKN2*.
(DOC)

Table S12 Probes and Primers used for TaqMan assays.
(DOC)

Table S13 Primers used for DNA re-sequencing.
(DOC)

Table S14 Primers used for construction of expression vectors.
(DOC)

Table S15 Oligonucleotides used for EMSAs and Luciferase assays.
(DOC)

Acknowledgments

We thank K. Kobayashi, M. Kitazato, K. Shimane, and all other members of the Laboratory for Autoimmune Diseases, CGM, RIKEN, for their advice and technical assistance. We also thank the members of BioBank Japan, the Rotary Club of Osaka-Midosuji District 2660 Rotary International, and Dr. Miyatake for supporting sample collection. The replication study of RA was performed under the support of the Genetics and Allied research in Rheumatic diseases Networking (GARNET) consortium.

Author Contributions

Conceived and designed the experiments: K Myouzen, Y Kochi, Y Okada, C Terao, K Ikari, K Ohmura, R Yamada, K Yamamoto. Performed the experiments: K Myouzen, Y Kochi, C Terao, A Suzuki, K Ikari, K Ohmura. Analyzed the data: K Myouzen, Y Kochi, Y Okada, C Terao, T Tsunoda, A Takahashi, R Yamada. Contributed reagents/materials/analysis tools: M Kubo, A Taniguchi, F Matsuda, K Ohmura, S Momohara, T Mimori, H Yamanaka, N Kamatani, Y Nakamura. Wrote the paper: K Myouzen, Y Kochi, Y Okada, C Terao, K Yamamoto.

References

1. Gabriel SE (2001) The epidemiology of rheumatoid arthritis. Rheum Dis Clin North Am 27: 269–281
2. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34: 395–402
3. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med 357: 1199–1209
4. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678
5. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, et al. (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 41: 820–823
6. 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
7. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 75: 330–337
8. Adrianto I, Wen F, Templeton A, Wiley G, King JB, et al. (2011) Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat Genet 43: 253–258
9. Thomas PD, Kejariwal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci U S A 101: 15398–15403
10. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, et al. (2012) A Genome-Wide Association Study Identified AFF1 as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese. PLoS Genet 8: e1002455. doi:10.1371/journal.pgen.1002455
11. 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
12. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073
13. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39: 1477–1482
14. 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
15. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, et al. (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 45: 511–516
16. Li Z, Nabel GJ (1997) A new member of the I kappaB protein family, I kappaB epsilon, inhibits RelA (p65)-mediated NF-kappaB transcription. Mol Cell Biol 17: 6184–6190
17. Whiteside ST, Epinat JC, Rice NR, Israel A (1997) I kappa B epsilon, a novel member of the I kappa B family, controls RelA and cRel NF-kappa B activity. Embo J 16: 1413–1426
18. Collier FM, Gregorio-King CC, Gough TJ, Talbot CD, Walder K, et al. (2004) Identification and characterization of a lymphocytic Rho-GTPase effector: rhotekin-2. Biochem Biophys Res Commun 324: 1360–1369
19. Collier FM, Loving A, Baker A, J., McLeod J, Walder K, et al. (2009) RTKN2 Induces NF-kappaB Dependent Resistance to Intrinsic Apoptosis in HEK cells and Regulates BCL-2 Gene in Human CD4+ Lymphocytes. J Cell Death 2: 9–23
20. Makarov SS (2001) NF-kappa B in rheumatoid arthritis: a pivotal regulator of inflammation, hyperplasia, and tissue destruction. Arthritis Res 3: 200–206
21. Kolbe D, Taylor J, Elnitski L, Eswara P, Li J, et al. (2004) Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. Genome Res 14: 700–707
22. Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, et al. (2006) ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 16: 1596–1604
23. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497–1502
24. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5: 829–834
25. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560
26. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43–49
27. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat Methods 3: 511–518
28. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246–1250
29. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26: 2474–2476
30. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315: 848–853
31. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224
32. 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
33. Chu X, Pan CM, Zhao SX, Liang J, Gao GQ, et al. (2011) A genome-wide association study identifies two new risk loci for Graves' disease. Nat Genet 43: 897–901
34. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, et al. (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet 43: 1193–1201
35. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR (2011) Low-coverage sequencing: implications for design of complex trait association studies. Genome Res 21: 940–951
36. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, et al. (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482: 390–394
37. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, et al. (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet 40: 1062–1064
38. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, et al. (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 40: 1216–1223
39. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324
40. 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
41. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10: 387–406

42. 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

43. Andrews NC, Faller DV (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res 19: 2499

Three Groups in the 28 Joints for Rheumatoid Arthritis Synovitis – Analysis Using More than 17,000 Assessments in the KURAMA Database

Chikashi Terao[1,2]*, Motomu Hashimoto[2,3], Keiichi Yamamoto[4], Kosaku Murakami[2], Koichiro Ohmura[2], Ran Nakashima[2], Noriyuki Yamakawa[2], Hajime Yoshifuji[2], Naoichiro Yukawa[2], Daisuke Kawabata[2], Takashi Usui[2], Hiroyuki Yoshitomi[5], Moritoshi Furu[3,5], Ryo Yamada[1,6], Fumihiko Matsuda[1,7,8], Hiromu Ito[3,5], Takao Fujii[2,3], Tsuneyo Mimori[2,3]

1 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan, **3** Department of the Control for Rheumatic Diseases, Kyoto University Graduate School of Medicine, Kyoto, Japan, **4** Department of Clinical Trial Design and Management, Translational Research Center, Kyoto University Hospital, Kyoto, Japan, **5** Department of Orthopaedic Surgery, Kyoto University Graduate School of Medicine, Kyoto, Japan, **6** Unit of Statistical Genetics Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **7** Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan, **8** CREST Program, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan

Abstract

Rheumatoid arthritis (RA) is a joint-destructive autoimmune disease. Three composite indices evaluating the same 28 joints are commonly used for the evaluation of RA activity. However, the relationship between, and the frequency of, the joint involvements are still not fully understood. Here, we obtained and analyzed 17,311 assessments for 28 joints in 1,314 patients with RA from 2005 to 2011 from electronic clinical chart templates stored in the KURAMA (Kyoto University Rheumatoid Arthritis Management Alliance) database. Affected rates for swelling and tenderness were assessed for each of the 28 joints and compared between two different sets of RA patients. Correlations of joint symptoms were analyzed for swellings and tenderness using kappa coefficient and eigen vectors by principal component analysis. As a result, we found that joint affected rates greatly varied from joint to joint both for tenderness and swelling for the two sets. Right wrist joint is the most affected joint of the 28 joints. Tenderness and swellings are well correlated in the same joints except for the shoulder joints. Patients with RA tended to demonstrate right-dominant joint involvement and joint destruction. We also found that RA synovitis could be classified into three categories of joints in the correlation analyses: large joints with wrist joints, PIP joints, and MCP joints. Clustering analysis based on distribution of synovitis revealed that patients with RA could be classified into six subgroups. We confirmed the symmetric joint involvement in RA. Our results suggested that RA synovitis can be classified into subgroups and that several different mechanisms may underlie the pathophysiology in RA synovitis.

Citation: Terao C, Hashimoto M, Yamamoto K, Murakami K, Ohmura K, et al. (2013) Three Groups in the 28 Joints for Rheumatoid Arthritis Synovitis – Analysis Using More than 17,000 Assessments in the KURAMA Database. PLoS ONE 8(3): e59341. doi:10.1371/journal.pone.0059341

Editor: Bernhard Kaltenboeck, Auburn University, United States of America

Received October 21, 2012; **Accepted** February 12, 2013; **Published** March 12, 2013

Funding: This study was supported by research grants from Mitsubishi Tanabe Pharma Corporation (http://www.mt-pharma.co.jp/e/), Eisai Co., Ltd. (http://www.eisai.co.jp/index.html), Abbott Japan Co., Ltd. (http://www.abbott.co.jp/), Chugai Pharmaceutical Co., Ltd. (http://www.chugai-pharm.co.jp/hc/ss/english/index.html), Pfizer Japan Inc. (http://www.pfizer.co.jp/pfizer/english/company/), and Bristol-Myyers K.K. (http://www.bms.co.jp/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The KURAMA database was supported by funding from Mitsubishi Tanabe Pharma Corporation, Eisai Co., Ltd., Abbott Japan Co., Ltd., Chugai Pharmaceutical Co., Ltd., Pfizer Japan Inc. and Bristol-Myyers. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: a0001101@kuhp.kyoto-u.ac.jp

Introduction

Rheumatoid arthritis (RA) is the most frequent inflammatory arthritis worldwide affecting 0.5 to 1% of the population [1]. As RA is a bone-destructive disease and functional impairment caused by joint damage is well correlated with swelling and tenderness of joints [2–3], the evaluation of joints in patients with RA is very important to assess disease activity and predict the risk of future joint deformity. ACR core set [4] and DAS (disease activity score) [5–6] were developed for evaluation of disease activity in RA. Recently, the three composite indices, namely, DAS28 [5], simplified disease activity index (SDAI) [7] and clinical disease activity index (CDAI) [8] are frequently used for disease activity evaluation among rheumatologists. All of the three indices are shown to be well correlated with future joint destruction [7,9]. These three methods include the same 28 joints for evaluation of disease activity, namely, bilateral wrist, 1st to 5th metacarpal (MCP) joints and proximal interphalangeal (PIP) joints, elbow, shoulder, and knee joints. Though RA is known to show symmetric joint symptoms [10], the frequency of bilateral joint symptoms and the correlations between each joint symptom are not fully analyzed by using large numbers of joint assessments. There are several reports of successful prediction of joint damage using a reduced number of joints for evaluation by ultrasonogra-

phy [11–12]. These reports raise the possibility that some of the 28 joints are less frequently involved, and are less informative for disease activity. Analyses for characterization of joint symptoms would uncover correlations of unexpected joint symptoms and distribution of synovitis in RA.

Here, we analyzed the distribution of affected joints in the 28 joints in patients with RA using more than 17,000 joint assessments from 1,314 patients with RA and showed that synovitis in RA patients can be classified into three groups. We also showed that affected rates of the 28 joints greatly vary in RA patients, and that RA patients could be classified into subgroups based on the distribution of joint synovitis.

Results

Frequency order of joints involvement

We recruited 17,311 assessments for the 28 joints in 1,314 patients with RA from 2005 to 2011. A summary of the registered patients is listed in Table 1. The distribution of the number of patients with RA in each year and the number of joint assessments for each patient are shown in Figure S1. We analyzed how often each of the 28 joints was tender or swollen in patients with RA in 2011. From the analysis of 735 patients, we found that the frequency of joint swelling and tenderness in the 28 joints is widely different from joint to joint (Figure 1 and Table S1). The wrist joints were the most frequently affected joints for swelling and tenderness. The frequency of the right wrist joint being affected was more than four times as high as the least frequently affected joint. Many of the joints showed right-dominant tenderness (eleven of fourteen joints, $p = 0.057$, binomial test), indicating mostly right-handedness. We found strong correlations for the affected rates of each joint between swellings and tenderness except for shoulder joints (Spearman's rank-sum coefficient, rho = 0.70 and $p = 3.8\times10^{-5}$, Figure 1, Table S1). Shoulder joints showed much higher frequencies of tenderness than those of swellings.

Next, we tried to replicate the order of affected frequencies of the 28 joints and the correlation between tenderness and swellings in different RA patients. We obtained 579 patients whose joints data were not available for 2011, indicating we analyzed independent RA patients. We found that the order of the affected joint frequencies were well correlated for both swelling and tenderness among different sets of RA patients (Spearman's rank-sum coefficient, rho:0.815 and 0.904, $p = 1.3\times10^{-7}$ and $p = 4.6\times10^{-11}$ for swelling and tenderness, respectively, Figure S2). We also confirmed that rates of tenderness were well correlated with those of swellings in the 28 joints in the 579 patients (rho:0.604). These results indicate that some of the 28 joints are more likely to develop arthritis than the others in RA patients. The swelling and tenderness correlate with each other except for shoulder joints.

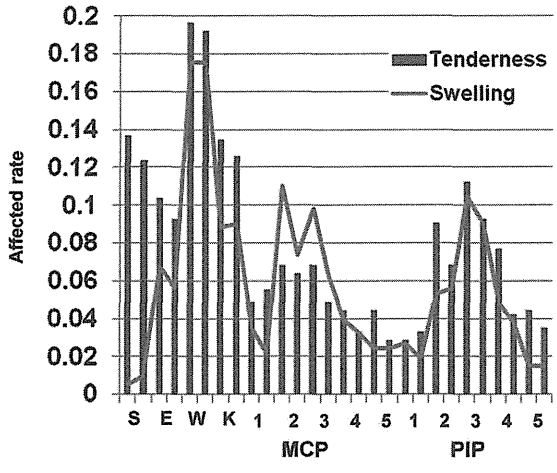


Figure 1. Affected rate of joint symptoms. Affected rate of joint symptoms. Each joint is arranged in the order of right and left. S:shoulder, E:elbow, W:wrist, K:knee.
doi:10.1371/journal.pone.0059341.g001

Table 1. Summary of the KURAMA database.

	The KURAMA database
RA patients	1314
Age (mean±SD)	60.2±15.1
female ratio	81.70%
disease duration (years)	12.2±9.8
Stage*	2.75±1.17
Class*	1.87±0.69

*Stage and Class represent Steinbrocker's stage and class, respectively.
SD: standard deviation.
doi:10.1371/journal.pone.0059341.t001

Whether the right-dominant involvement of joints in patients with RA is associated with joint destruction was analyzed. Joint destruction in the hand was evaluated for 246 patients with RA by modified Sharp score [13]. The six elements of the scores were separately analyzed, namely erosion of PIP, MCP, and wrist joints (we defined as joints other than MCP and PIP in hand) and narrowing of PIP, MCP, and wrist joints. We found that five out of six elements showed right-dominant destruction. In particular, narrowing and erosion of MCP joints showed a statistically significant right-dominance in binomial test ($p <= 0.0050$, Table S2).

Three groups of 28 joints in RA synovitis

Next we analyzed correlations of joint symptoms between the 28 joints. We randomly picked up one assessment from each of the 1,314 patients to maximize the power. When the correlation of tenderness of the 28 joints was analyzed with kappa coefficient, we confirmed that each joint showed a symmetric involvement (Figure 2A). The results also showed that the tenderness of large joints and wrist joints are not correlated with the tenderness of PIP and MCP joints. We found that the tenderness of MCP joints was especially well correlated with each other and that PIP joints tenderness was well correlated with each other. The correlation of swelling in the 28 joints showed the same tendency as that of tenderness, namely, symmetric joint involvement, correlations between large joints and wrist joints, and no strong correlations between wrist joints and other small joints (Figure 2B).

Next we used eigen vectors of principal component analysis to assess the correlations of the 28 joints involvement. When we analyzed correlations of tenderness, eigen vectors revealed that PIP and MCP joints can be clearly distinguished from large joints and wrist joints (Figure 3A). PIP joints and MCP joints turned out to make independent groups after excluding large joints and wrist joints (Figure 3B). These three groups of affected joints were found both for tenderness and swelling (Figure 3C and 3D). We confirmed these three correlation groups in four independent resampling analyses by randomly picking up one assessment from each of the 1,314 patients four times (data not shown). The three groups were observed in the two independent sets of RA patients which were used in the analysis of joints involvement frequency

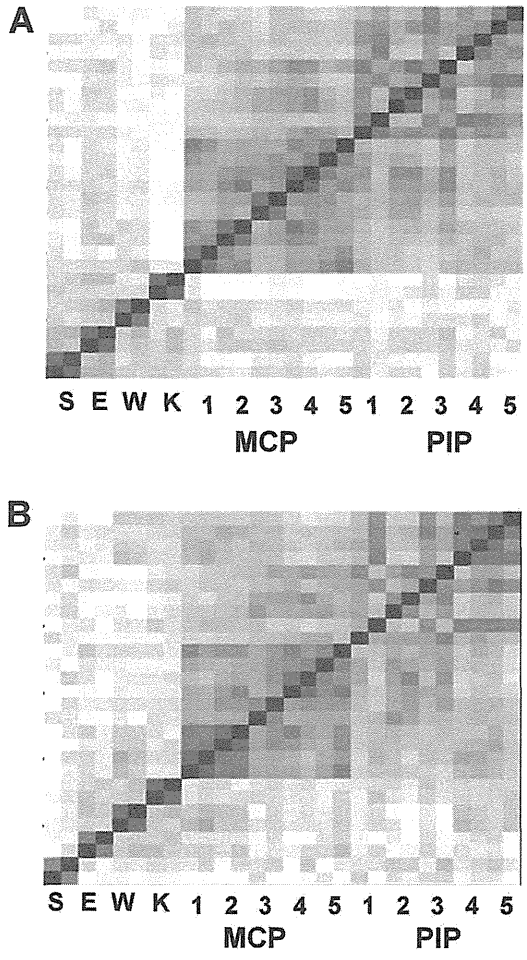


Figure 2. Correlations between the 28 joint symptoms. Brightness of the red color corresponds to the strength of correlations between joint tenderness (A) or swellings (B), using the Kappa coefficient. Each joint is arranged in the order of right and left. The joint order in the y axis is the same as the x axis. The result is a representative of five analyses based on resampled assessments. S:shoulder, E:elbow, W:wrist, K:knee.
doi:10.1371/journal.pone.0059341.g002

(Figure S3). In addition, no significant difference was observed in the relationship of the three groups of joint involvement when we divided the 1,314 patients into two groups according to the patients' caring physicians (Figure S4). We confirmed the three groups by resampling four times for each analysis (data not shown). These results indicate that these three groups were not due to specific patients, examiners, or time of evaluation.

Taken together, the correlation analyses using kappa coefficient and eigen vectors in principal component analysis indicated that there are three correlated groups of joints in RA synovitis, namely, large joints with wrist joints (which we express as "large and wrist joints"), PIP joints, and MCP joints.

Subgroups of patients with RA

We performed a clustering analysis of 5,383 evaluations of 28 joints from 1,314 patients with RA. Six subgroups of evaluations of 28 joints were observed (Figure 4). Each of the subgroups was characterized by 1) no synovitis (34.6%), 2) mild activity with dominant involvement of large and wrist joints (17.4%), 3) dominant involvement of MCP joints (18.3%), 4) dominant involvement of PIP joints (9.3%), 5) active synovitis (4.1%), and 6) moderate activity with dominant involvement of large and wrist joints (16.4%) (Table S3). Whether patients with RA are classified into the same subgroups was analyzed. There were 998 patients with four or five evaluations, and of these, 734 were categorized into the regular groups across different evaluations, indicating that the patterns of synovitis in the same patients were stable. Analysis of joint destruction in each subgroup revealed that the sixth subgroup demonstrated dominant destruction of large and wrist joints compared with MCP and PIP joints ($p <= 2.8 \times 10^{-5}$, Figure S5 and Figure S6).

Discussion

Since RA is a joint destructive autoimmune arthritis and joint damage occurs rapidly in the early stages of the disease course [14], the development of a quantitative scale which assesses disease activity and predicts joint damage is very important. After DAS and ACR core sets were introduced, DAS28, SDAI, and CDAI were developed to evaluate disease activity and easily calculate the disease activity score in patients with RA. All three indices were shown to be well correlated with future joint destruction and they share the same 28 joints for evaluation. Joint symptoms especially joint swelling is known to correlate with future joint damage [3]. While these indices were developed for use in clinical trials such as responsiveness to treatment, they are used by rheumatologists in daily clinical practice and they are reported to coincide very well among different examiners [9]. Characterizing the relative affected frequency of each joint and analysis of correlation between joint symptoms are important to analyze the basic mechanisms of synovitis and to efficiently select the joints to predict future joint destruction. However, there is no detailed analysis to address the correlations between the 28-joint symptoms.

In the current study, we characterized the 28-joint symptoms using large numbers of joint assessments. While we reported the affected rates of each joint in the 28 joints for tenderness and swelling of RA patients registered in the KURAMA database in 2011 as a representative (Table S1), these rates should not be generalized considering large effects of treatment especially biologics agents on joint symptoms. Thus, we focused on relative frequencies of joint involvement for the 28 joints. The affected frequency pattern was compared between the two sets of RA patients, and there were no apparent differences between the two sets for both tenderness and swelling. We also showed that joint symptoms in RA could be classified into three groups both for tenderness and swelling. Our analysis also demonstrated that patients with RA can be regularly classified into six subgroups based on patterns of joint symptoms. These results suggest that regular RA joint involvement pattern, including relative frequency and groups of joints, is largely maintained in RA patients. In addition, we confirmed that these patterns of joint involvement were not attributed to evaluators and fractions of RA patients.

It is interesting that the affected frequencies greatly varied from joint to joint, and the rate of the most highly affected joint was more than four times as high as the least-affected joint. The affected frequencies indicated that wrist joints were the most frequently affected. It should be noted that surface area may have influenced the sensitivity of detecting synovitis in physical exams when different joints were compared. The relatively high frequency of tenderness and swelling in large and wrist joints compared with MCP and PIP joints can be explained by this difference in surface area. However, surface area cannot fully explain the highest frequency of wrist involvement and different frequencies within the MCP or PIP joints. A dominant involve-

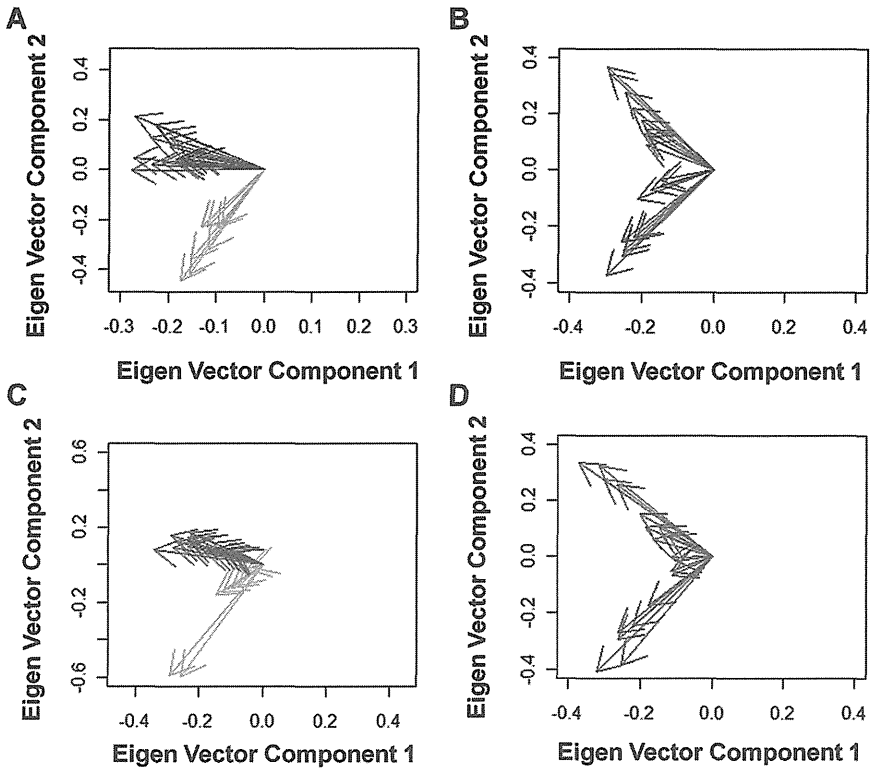


Figure 3. Relationship of the 28-joint involvement. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). The results are representatives of five analyses based on resampled assessments. Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
doi:10.1371/journal.pone.0059341.g003

ment of right joints seemed to indicate a majority of the study population being right-handed in spite of the small difference of affected rates between bilateral joints. We also demonstrated that the right dominant involvement was also true for joint destruction. We could not compare the joint involvement and joint destruction between right-handed patients and left-handed patients due to a lack of information regarding handedness of patients.

Correlation analysis confirmed the well-known symmetric joint involvement in patients with RA. Strong correlations of tenderness and swelling in the same joints except for shoulder joints may indicate low sensitivity of shoulder swelling in the physical exams and common mechanisms of swelling and tenderness. It is striking that joint symptoms can be classified into three groups based on correlation analysis and principal component analysis. The association observed between the symptoms in the wrist joints and the large joints is worth noting, since wrist joints are regarded as small joints according to ACR/EULAR criteria set in 2010. As wrist joints are much closer to other small joints than large joints, the relationship between wrist joints and large joints cannot be explained by the distance of joints. The distance of joints cannot explain the two different groups of MCP and PIP joints either. While symptoms of large and wrist joints are not related with those of MCP and PIP joints, they were not very strongly correlated with each other, compared with correlations among PIP joints or MCP joints. This may indicate that there are no common strong factors which predispose large and wrist joints to swelling and tenderness in patients with RA.

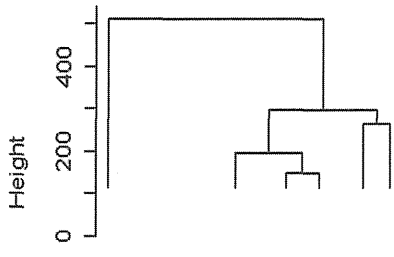


Figure 4. Six subgroups of evaluations of the 28 joints in RA. Results of clustering analysis with Ward method using randomly obtained 5,383 evaluations of the 28 joints in 1,314 patients were plotted.
doi:10.1371/journal.pone.0059341.g004

We also showed that patients with RA can be divided into six subgroups based on these three groups of joint involvement. More than 70% of patients are classified into regular subgroups, indicating that the pattern of synovitis in a patient with RA is stable. When patients who were regularly classified into the first subgroup of patients characterized by no synovitis were removed, more than 60% of patients were still classified into regular subgroups (data not shown), indicating that the stable patterns were observed regardless of activity of RA. As joint destruction was influenced by disease duration, disease activity, and treatment, we analyzed the relative distribution of joint destruction between the three joint groups in a patient with RA. We found that the sixth subgroup of patients, characterized by moderate activity with dominant involvement of large and wrist joints, demonstrated dominant destruction of wrist joints. This suggests that classifying patients with RA into appropriate subgroups would lead to prediction of patterns of joint destruction.

There are reports that evaluating fraction of joints by ultrasonography is a good way to predict future joint damage [11–12]. One study reported that 5 of the 28 joints with MTP2 and MTP5 joints, namely, wrist, MCP2, MCP3, PIP2, and PIP3 joints, are enough for ultrasonography evaluation [12]. Their data seems to be consistent with our results as they selected at least two joints from three different groups into which the 28-joint symptoms were classified. As ultrasonography usually surpasses physical examination in terms of the sensitivity to detect synovitis, it is interesting to analyze whether the assessments of synovitis using ultrasonography show the same pattern of synovitis over the 28 joints in RA.

Our results indicate that RA does not develop synovitis in the 28 joints with the same frequency and that the affected rate of each joint greatly varies from joint to joint. These different distributions of joint synovitis would lead to different distribution of joint destruction. Based on our results, the 28 joints can be categorized into three groups, and it is possible that some fractions of the 28 joints are less informative to assess disease activity than others. It would be interesting to develop a novel simplified joint core set, and analyze the correlation between joint damage and activity score based on this. It would be also interesting to characterize each of RA subsets in more detail.

Materials and Methods

Ethics Statement

Written informed consent to enroll in the database described below was obtained from most of the patients, but for some patients the information regarding the construction of this database was disclosed instead of obtaining written informed consent. Participants who were informed regarding the construction of the database (instead of obtaining written informed consent) were allowed to withdraw from the study if desired.

All data were de-identified and analyzed anonymously. This study was designed in accordance with the Helsinki Declaration. This study including the consent procedure was approved by the ethics committee of Kyoto University Graduate School and Faculty of Medicine.

The KURAMA database

The KURAMA (Kyoto University Rheumatoid Arthritis Management Alliance) database was established in 2011 at Kyoto University to store detailed clinical information and specimens from patients with arthritis and arthropathy. The alliance is composed of rheumatic disease-associated departments in Kyoto University Hospital as well as its allied, integrating previous database and specimen collections in each department and allied. A template for electronic clinical charts developed at Kyoto University Hospital in 2004 to evaluate joint involvements in RA patients was used to obtain joint assessments. Rheumatologists evaluated swelling and tenderness of the 28 joints in patients with RA on each visit and filled in the template. The synovitis information of the 28 joints and data for C-reactive protein and erythrocyte sedimentation rate were extracted from electronic clinical charts [15] and stored in the KURAMA database.

Patients and data of joint assessment

A total of 17,311 joint assessments from 1,314 patients with RA from 2005 to 2011 were obtained in a retrospective manner from the KURAMA database. All of the patients fulfilled ACR revised criteria for RA in 1987 [10] or ACR and EULAR classification criteria for RA in 2010 [16–17].

Analysis of affected frequencies in the 28 joints

RA patients were subdivided depending on whether their data were available in 2011 or not, and the affected frequency in each of the 28 joints was calculated. We compared the order of the affected frequency in the 28 joints between the two patient sets with Spearman's rank-sum coefficient. We separately analyzed the affected rates of joints for swelling and tenderness. When multiple joint assessments in different visits were available in the same patient with RA, we randomly selected one of the assessments as representative in the patient. We compared frequencies between tenderness and swellings for the 28 joints with Spearman's rank-sum coefficient.

Clustering of patients with RA

Clustering analyses were performed by Ward method, using randomly-selected 5,383 evaluations of the 28 joints from 1,314 patients with RA. These evaluations did not contain more than six assessments from each patient to avoid excess influence of particular patients. Affected rates were calculated for the three groups of joints (namely PIP joints, MCP joints and large and wrist joints) in this clustering analysis. For example, when a patient showed tenderness and swelling for all PIP joints, the affected rate of PIP joints in the patient is 2. When a patient showed tenderness for four MCP joints, the affected rate of MCP joints is 0.4.

RA patients were regarded as belonging to a particular group when more than 60% of evaluations belonging to the same patients with four or five evaluations were classified into the same group.

Analysis between RA subgroups and joint destruction

Joint destruction of hand joints in 246 patients with RA was evaluated by modified Sharp score by a trained rheumatologist who was not informed of the patients' characteristics (KM). Joint destruction rates were defined for the three groups of joints as a sum of scores divided by the full score in the joints group. For example, when a patient shows 50 as a sum of scores in the large and wrist group, the patient's joint destruction rate for the group is 0.463 (50/108).

Correlation of the 28 joints and statistical analysis

Correlations of joint symptoms among the 28 joints were estimated separately for tenderness and swelling. We randomly obtained one assessment of the 28 joints in each patient as a representative of the patient's joint assessments for maximization of the power. Kappa coefficient was used to analyze coincidence of joint symptoms in each pair of the 28 joints. Eigen vectors obtained in principal component analysis were used to analyze the deviation of joint symptoms. We resampled joint assessments for each patient and created four other sets of joint assessments. The same correlation analyses were performed using the four resampled assessments to confirm the correlation shown in the first assessment set. Right dominance of the synovitis and joint destruction was analyzed by binomial test. Dominant destruction of joints was evaluated by paired-t test. Statistical analysis was performed by R software or SPSS (ver18).

Supporting Information

Figure S1 Distribution of joint evaluation counts and patients across different years. A) Distribution of number of RA patients according to numbers of 28-joint assessments. B) Distribution of number of patients with RA whose joint assessment data were available from 2005 to 2011 in the KURAMA database. (TIF)

Figure S2 Good correlations between joint involvement rates in different sets of RA patients. Rates of joint involvement for A) swelling and B) tenderness were compared between the two different sets of RA patients. X and Y axes represent rates in the first set of RA patients in 2011 and those in the second set in 2005 to 2010, respectively.
(TIF)

Figure S3 Three groups of joints regardless of different sets of RA patients. Analysis using one of four resampled assessments in one of the two sets of RA patients is shown as a representative. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
(TIF)

Figure S4 Three groups of joints regardless of different evaluators. Analysis using one of five resampled assessments by one of the two groups of medical doctors is shown as a representative. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
(TIF)

Figure S5 Dominant destruction of large and wrist joints in the sixth subgroup of patients with RA. Box plots indicating the joint destruction rates in the three joint groups in subjects belonging to the sixth subgroup.
(TIF)

Figure S6 Destruction of large and wrist joints among the six subgroups of RA. Differences in destruction rates were plotted for each subject in the six subgroups. The difference was defined as: A) destruction rate of group of large and wrist joints – destruction rate of MCP joints and B) destruction rate of group of large and wrist joints – destruction rate of PIP joints.
(TIF)

Table S1 Rate of joint involvement for 28 joints in RA.
(DOC)

Table S2 Right-dominant joint destruction in RA. Patients who showed unilateral higher or lower scores in each element were analyzed.
(DOC)

Table S3 Mean affected rates of the three joint groups in the six subgroups of patients with RA.
(DOC)

Acknowledgments

We would like to thank to Mr. Wataru Yamamoto at Kurashiki Kosai Hospital for his excellent support to establish and maintain the KURAMA database. We also thank Drs Hisashi Yamanaka, Katsunori Ikari, and Ayako Nakajima at Institute of Rheumatology, Tokyo Women's Medical University for their kind instruction and advice for management of rheumatic diseases database.

Author Contributions

Evaluation of joint X-rays: KM. Conceived and designed the experiments: CT MH KO RY FM HI TF TM. Analyzed the data: CT. Contributed reagents/materials/analysis tools: CT MH KO RN KM N. Yamakawa H. Yoshifuji N. Yukawa DK TU H. Yoshitomi MF HI TF TM KY. Wrote the paper: CT.

References

1. Firestein GS (2003) Evolving concepts of rheumatoid arthritis. Nature 423: 356–361.
2. Drossaers-Bakker KW, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, et al. (1999) Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. Arthritis and Rheumatism 42: 1854–1860.
3. Smolen JS, Van Der Heijde DM, St Clair EW, Emery P, Bathon JM, et al. (2006) Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial. Arthritis and Rheumatism 54: 702–710.
4. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, et al. (1993) The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis and Rheumatism 36: 729–740.
5. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, et al. (1990) Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Annals of the Rheumatic Diseases 49: 916–920.
6. van der Heijde DM, van't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, et al. (1992) Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. Annals of the Rheumatic Diseases 51: 177–181.
7. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, et al. (2003) A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology 42: 244–257.
8. Aletaha D, Smolen JS (2007) The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract Res Clin Rheumatol 21: 663–675.
9. Salaffi F, Cimmino MA, Leardini G, Gasparini S, Grassi W (2009) Disease activity assessment of rheumatoid arthritis in daily practice: validity, internal consistency, reliability and congruency of the Disease Activity Score including 28 joints (DAS28) compared with the Clinical Disease Activity Index (CDAI). Clinical and Experimental Rheumatology 27: 552–559.
10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324.
11. Scheel AK, Hermann KG, Kahler E, Pasewaldt D, Fritz J, et al. (2005) A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis. Arthritis and Rheumatism 52: 733–743.
12. Backhaus M, Ohrndorf S, Kellner H, Strunk J, Backhaus TM, et al. (2009) Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project. Arthritis and Rheumatism 61: 1194–1201.
13. van der Heijde D (2000) How to read radiographs according to the Sharp/van der Heijde method. Journal of Rheumatology 27: 261–263.
14. Machold KP, Stamm TA, Eberl GJ, Nell VK, Dunky A, et al. (2002) Very recent onset arthritis – clinical, laboratory, and radiological findings during the first year of disease. Journal of Rheumatology 29: 2278–2287.
15. Yamamoto K, Yamanaka K, Hatano E, Sumi E, Ishii T, et al. (2012) An eClinical trial system for cancer that integrates with clinical pathways and electronic medical records. Clin Trials 9: 408–417.
16. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, et al. (2010) 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis and Rheumatism 62: 2569–2581.
17. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, et al. (2010) 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Annals of the Rheumatic Diseases 69: 1580–1588.

Genome-Wide Association Study and Gene Expression Analysis Identifies *CD84* as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis

**Jing Cui[1☯], Eli A. Stahl[1,2,3☯¤a], Saedis Saevarsdottir[4,5], Corinne Miceli[6,7], Dorothee Diogo[1,2,3], Gosia Trynka[1,2,3], Towfique Raj[2,3,8], Maša Umićević Mirkov[9], Helena Canhao[1,10,11], Katsunori Ikari[12], Chikashi Terao[13,14], Yukinori Okada[1,2,3], Sara Wedrén[4,5], Johan Askling[4,15], Hisashi Yamanaka[12], Shigeki Momohara[12], Atsuo Taniguchi[12], Koichiro Ohmura[13], Fumihiko Matsuda[14], Tsuneyo Mimori[13], Namrata Gupta[3], Manik Kuchroo[3,8], Ann W. Morgan[16], John D. Isaacs[17], Anthony G. Wilson[18], Kimme L. Hyrich[19], Marieke Herenius[20], Marieke E. Doorenspleet[20], Paul-Peter Tak[20¤b], J. Bart A. Crusius[21], Irene E. van der Horst-Bruinsma[22], Gert Jan Wolbink[23,24,25], Piet L. C. M. van Riel[9], Mart van de Laar[26], Henk-Jan Guchelaar[27], Nancy A. Shadick[1], Cornelia F. Allaart[28], Tom W. J. Huizinga[28], Rene E. M. Toes[28], Robert P. Kimberly[29], S. Louis Bridges Jr.[29], Lindsey A. Criswell[30], Larry W. Moreland[31], João Eurico Fonseca[10,11], Niek de Vries[20], Barbara E. Stranger[2,3], Philip L. De Jager[2,3,7], Soumya Raychaudhuri[1,2,3,32], Michael E. Weinblatt[1], Peter K. Gregersen[33], Xavier Mariette[6,7], Anne Barton[34], Leonid Padyukov[5], Marieke J. H. Coenen[9], Elizabeth W. Karlson[1], Robert M. Plenge[1,2,3]*

1 Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Medical and Population Genetics Program, Chemical Biology Program, Broad Institute, Cambridge, Massachusetts, United States of America, **4** Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden, **5** Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, **6** Université Paris-Sud, Orsay, France, **7** APHP–Hôpital Bicêtre, INSERM U1012, Le Kremlin Bicêtre, Paris, France, **8** Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **9** Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **10** Rheumatology Research Unit, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal, **11** Rheumatology Department, Santa Maria Hospital–CHLN, Lisbon, Portugal, **12** Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, **13** Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan, **14** Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **15** Clinical Epidemiology Unit, Department of Medicine, Karolinska Institute/Karolinska University Hospital, Stockholm, Sweden, **16** NIHR–Leeds Musculoskeletal Biomedical Research Unit and Leeds Institute of Molecular Medicine, University of Leeds, Leeds, United Kingdom, **17** Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle Upon Tyne, United Kingdom, **18** Rheumatology Unit, Medical School, University of Sheffield, Sheffield, United Kingdom, **19** School of Translational Medicine, Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester, United Kingdom, **20** Department of Clinical Immunology and Rheumatology, Academic Medical Center/University of Amsterdam, Amsterdam, The Netherlands, **21** Laboratory of Immunogenetics, Department of Pathology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands, **22** Department of Rheumatology, Vrije Universiteit University Medical Center, Amsterdam, The Netherlands, **23** Sanquin Research Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **24** School of Medicine and Biomedical Sciences, Sheffield University, Sheffield, United Kingdom, **25** Jan van Breemen Institute, Amsterdam, The Netherlands, **26** Arthritis Center Twente, University Twente and Medisch Spectrum Twente, Enschede, The Netherlands, **27** Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, The Netherlands, **28** Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands, **29** Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **30** Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, California, United States of America, **31** Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **32** NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, **33** The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, United States of America, **34** Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom

Abstract

Anti-tumor necrosis factor alpha (anti-TNF) biologic therapy is a widely used treatment for rheumatoid arthritis (RA). It is unknown why some RA patients fail to respond adequately to anti-TNF therapy, which limits the development of clinical biomarkers to predict response or new drugs to target refractory cases. To understand the biological basis of response to anti-TNF therapy, we conducted a genome-wide association study (GWAS) meta-analysis of more than 2 million common variants in 2,706 RA patients from 13 different collections. Patients were treated with one of three anti-TNF medications: etanercept (n = 733), infliximab (n = 894), or adalimumab (n = 1,071). We identified a SNP (rs6427528) at the *1q23* locus that was associated with change in disease activity score (ΔDAS) in the etanercept subset of patients ($P = 8 \times 10^{-8}$), but not in the infliximab or adalimumab subsets ($P > 0.05$). The SNP is predicted to disrupt transcription factor binding site motifs in the 3′ UTR of an immune-related gene, *CD84*, and the allele associated with better response to etanercept was associated with higher *CD84* gene expression in peripheral blood mononuclear cells ($P = 1 \times 10^{-11}$ in 228 non-RA patients and $P = 0.004$ in 132 RA patients). Consistent with the genetic findings, higher *CD84* gene expression correlated with lower cross-sectional DAS ($P = 0.02$, n = 210) and showed a non-significant trend for better ΔDAS in a subset of RA patients with gene expression data (n = 31, etanercept-treated). A small, multi-ethnic replication showed a non-significant trend towards an association among etanercept-treated RA

patients of Portuguese ancestry (n = 139, $P = 0.4$), but no association among patients of Japanese ancestry (n = 151, $P = 0.8$). Our study demonstrates that an allele associated with response to etanercept therapy is also associated with *CD84* gene expression, and further that *CD84* expression correlates with disease activity. These findings support a model in which *CD84* genotypes and/or expression may serve as a useful biomarker for response to etanercept treatment in RA patients of European ancestry.

Citation: Cui J, Stahl EA, Saevarsdottir S, Miceli C, Diogo D, et al. (2013) Genome-Wide Association Study and Gene Expression Analysis Identifies *CD84* as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis. PLoS Genet 9(3): e1003394. doi:10.1371/journal.pgen.1003394

Editor: Alison Motsinger-Reif, North Carolina State University, United States of America

Received August 13, 2012; **Accepted** January 13, 2013; **Published** March 28, 2013

Funding: JC is supported by grants from ACR REF HPNIA award and the NIH (R01-AR059073, P60-AR047782, U01-GM092691, R01-AR049880). RMP is supported by grants from the NIH (R01-AR057108, R01-AR056768, U01-GM092691, R01-AR059648) and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. The American College of Rheumatology Research and Education Foundation provided funding to support this project. The eRA study was supported by R01 AI/AR47487. SS's work was supported by a clinical research fund from Stockholm County (ALF fund). LP's work was supported by the Swedish Rheumatism Association and the Swedish Medical Research Council. NdV and MED were sponsored by CTMM, the Center for Translational Molecular Medicine, and the Dutch Arthritis Foundation project TRACER (grant 04I-202). The Swedish studies (EIRA, Karolinska) were supported by grants from the Swedish Medical Research Council, the Stockholm County Council, the Swedish Council for Working Life and Social Research, King Gustaf V's 80-year foundation, the Swedish Rheumatism Association, the Swedish Foundation for Strategic Research, the Swedish COMBINE project, and the IMI funded BTCure project. 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: rplenge@partners.org

¤a Current address: Division of Psychiatric Genomics, Mt. Sinai School of Medicine, New York, New York, United States of America
¤b Current address: GlaxoSmithKline, Stevenage, United Kingdom

☯ These authors contributed equally to this work.

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation of the synovial lining of the joint [1]. If left untreated, outcome varies from self-limited disease in a small proportion of RA patients to severe disease resulting in profound structural damage, excess morbidity and disability, and early mortality [2]. In the last twenty years, disease activity has been controlled in many patients by treatment with disease-modifying anti-rheumatic drugs (DMARDs), such as methotrexate, and the more recently developed biologic DMARDs that block inflammatory cytokines such as tumor necrosis factor-alpha (TNFa) [3]. Unfortunately, these medications are not effective in all RA patients, with up to one-third of patients failing to respond to any single DMARD [1–3]. Moreover, the biological mechanisms underlying treatment failure are unknown, which limits the development of clinical biomarkers to guide DMARD therapy or the development of new drugs to target refractory cases.

There are two classes of anti-TNF therapy: the TNF receptor fusion protein (etanercept), which acts as a soluble receptor to bind circulating cytokine and prevent TNF from binding to its cell surface receptor, and monoclonal antibodies that bind TNF (adalimumab, infliximab, certolizumab, and golimumab). There are undoubtedly shared mechanisms between the two drug classes (e.g., downstream signaling factors), as illustrated by similar effects on the change in inflammatory cytokines, complement activation, lymphocyte trafficking, and apoptosis [4,5,6]. Similarly, there are likely to be different biological factors that influence response: infliximab and adalimumab are approved for treatment of Crohn's disease; infliximab and adalimumab bind to transmembrane TNF on the surface of activated immune cells, whereas etanercept only binds soluble TNF [7]; and etanercept also binds a related molecule, lymphotoxin alpha (LTA), whereas infliximab/adalimumab do not [8].

Pharmacogenetics of response to anti-TNF therapy in RA remains in its early stages, with no single variant reaching an unambiguous level of statistical significance. Candidate gene studies suggest associations of TNFa or TNF receptor alleles, RA risk alleles or other SNPs with response to anti-TNF therapy [9,10,11]. Two GWAS in small sample sets (largest was 566 patients) have been performed, which identified loci with suggestive evidence for association [12,13]. Therefore, GWAS of large sample sizes may yet uncover genetic factors associated with response to anti-TNF therapy in RA, and larger cohorts enable separate analyses of the different types of anti-TNF drugs.

Here we report a GWAS of 2,706 samples with anti-TNF treatment response data collected from an international collaboration, including previously published GWAS data [12,13]. Our primary outcome measure was the change in disease activity score based on a joint count in 28 joints (DAS28) from baseline to 3–12 months after initiating anti-TNF therapy. Our secondary outcome measure was European League Against Rheumatism (EULAR) responder status [14,15], where patients are classified as EULAR good responders, moderate responders or non-responders based on follow up DAS28 after treatment and overall change in DAS28. We found a highly significant association for a variant that we also show is also a strong expression quantitative trait locus (eQTL) for the *CD84* gene. Our findings suggest that *CD84* genotype and/or expression may prove to be a biomarker for etanercept response in RA patients.

Results

Genome-wide association study

Clinical and GWAS data were compiled for 2,706 individuals of European ancestry from 13 collections as part of an international collaboration. Table 1 shows sample sizes, phenotypes and clinical variables for the four collections that were the units of analysis (additional details are shown in Table S1). Disease activity score based on a 28-joint count (DAS28) were collected at baseline and at one time point after anti-TNF therapy administration (mean 3.7 months, range 3–12 months). We defined our primary phenotype

Author Summary

There are no genetic predictors of response to one of the most widely used classes of drugs in the treatment of rheumatoid arthritis—biological modifiers of the inflammatory cytokine tumor necrosis factor-alpha (or anti-TNF therapy). To identify genetic predictors, we performed the largest genome-wide association study (GWAS) to date as part of an international collaboration. In our study, which included 2,706 RA patients treated with one of three anti-TNF drugs, the most significant finding was restricted to RA patients treated with etanercept ($P=8\times10^{-8}$), a drug that acts as a soluble receptor to bind circulating cytokine and prevents TNF from binding to its cell surface receptor. The associated variant influences expression of a nearby immune-related gene, *CD84*, whose expression is correlated with disease activity in RA patients. Together, our data support a model in which genomic factors related to *CD84* expression serve as a predictor of disease activity and response to etanercept therapy among RA patients of European ancestry, but not anti-TNF therapies that act through different biological mechanisms or potentially in RA patients of other genetic ancestries.

as a change in DAS28 (ΔDAS) from baseline (so that greater ΔDAS corresponded with better response to therapy; overall mean and standard deviation of 2.1 ± 1.3), adjusted for baseline DAS. A secondary phenotype was used based on European League Against Rheumatism (EULAR) response criteria. EULAR 'good response' was defined as ending DAS$<$3.2 and ΔDAS$>$1.2; 'non-response' was defined as ΔDAS $<$0.6 or ΔDAS$\leq$1.2, and ending DAS $>$5.1; and 'moderate response' is in between [15]. We limited our secondary analysis to a dichotomous outcome, EULAR good responders (n = 998 for all patients treated with anti-TNF therapy) versus EULAR non-responders (n = 655), excluding the moderate category based on the hypothesis that a more extreme phenotype of response would yield improved discrimination.

Clinical variables were examined for association with phenotype, and therefore possible confounding in genetic association tests. In multivariate models (Table S2), only baseline DAS was strongly associated with the ΔDAS phenotype. As previously shown [11], age and gender showed univariate associations that were attenuated in the multivariate analysis. Accordingly, we used only baseline DAS as a clinical covariate, as this allowed us to maximize sample size given clinical variable missing data in some cohorts.

We performed quality control (QC) filtering and data processing of GWAS data for each of eleven genotyping batches. Genotyping array platforms are described in the Methods. HapMap2 imputation allowed us to test for association at $>$2 M SNPs with imputation quality scores $>$0.5. Genotype data were merged across several genotype batches to create four collections for genome-wide association testing. We performed linear regression association tests using baseline DAS and three principal components as covariates, and performed inverse-variance weighted meta-analysis to combine results across the four collections. Quantile-quantile plots with genomic control λ_{GC} values are shown in Figure S1. We found no evidence of systematic inflation of association test results, and no evidence of deflation for imputed versus genotyped SNPs. As a final filter, we excluded SNPs that

Table 1. Samples and clinical data.

Collection (analysis batch):	REF	BRAGGSS	DREAM	ReAct	Total
Sample sizes	959*	595	880*	272	2706
Drug subsets					
etanercept	365	259	109	0	733
infliximab	415	268	211	0	894
adalimumab	174	68	557	272	1071
EULAR Reponse categories					
Good responder	432**	161	313	92	998
Moderate responder	243	258	359	131	991
Non-responder	322	176	208	49	755
Genotype platform	mixed	Affy 500K	Illu550K +650K	Illumina OmniExpress	
Clinical variables					
Age, yr; mean (SD)	53.6 (12.7)	57.4 (10.9)	54.8 (12.9)	53.9 (10.8)	
Disease duration, yr; mean (SD)	6.7 (9.4)	14 (9.8)	9.6 (9.5)	12 (9.1)	
Gender, female %	75.6	77.3	68.3	77.9	
Seropositive, %	87	78	80	70	
MTX co-therapy, %	65.6	85.6	76.0	50.0	
Baseline DAS, mean (SD)	5.5 (1.2)	6.7 (0.9)	5.5 (1.2)	5.9 (1.0)	
ΔDAS, mean (SD)	1.9 (1.6)	2.5 (1.5)	1.9 (1.3)	2.2 (1.3)	
Mean treatment duration	4.6	5.6	3	3	
Study design	All***	Observational	Observational	Observational	

*8 patients had no TNF drug information.
**38 patients had only EULAR response (good, moderate or none) clinical data.
***ABCoN, GENRA are prospective cohorts, BeSt, eRA and TEAR are randomized controlled trial (RCT), and rest of REF group are observational cohorts.
doi:10.1371/journal.pgen.1003394.t001

showed strong evidence of heterogeneity across collections (Cochran's Q $P<0.001$).

We first analyzed all samples together (n = 2,706), regardless of drug type. We found no clear evidence of association with treatment response measured by ΔDAS (Figure 1A). Similar results were obtained using the binary phenotype of EULAR responder versus EULAR non-responder status (Figures S1 and S2).

We next separately analyzed patients treated with either etanercept (n = 733), infliximab (n = 894) or adalimumab (n = 1,071) (Figure 1B–1D), under the hypothesis that different genetic loci affect response to the different drugs based on their mechanism of action or other biochemical properties. GWAS results are publicly available for all SNPs tested at the Plenge laboratory and RICOPILI Web sites (see URLs). GWAS results for all SNPs achieving $P<10^{-6}$ from any analysis are detailed in the Table S3.

For etanercept-treated RA patients, a locus on chromosome *1q23* achieved near-genome-wide significance (rs6427528, $P_{META}=8\times10^{-8}$) (Figure 1B, Figure 2A, and Figure 3), but not in the infliximab or adalimumab subsets ($P>0.05$) (Figure S3). SNPs in linkage disequilibrium (LD) showed consistent association results (rs1503860, $P=1\times10^{-7}$, $r^2=1$ with rs6427528 in HapMap; three perfect-LD clusters of SNPs exemplified by rs3737792, rs10908787 and rs11265432 respectively; $P<5\times10^{-6}$; $r^2=0.83$, 0.63 and 0.59 with rs6427528, respectively). No single collection was responsible for the signal of association, as the effect size was consistent across all collections (Figure S4). The top SNP rs6427528 was genotyped in the ReAct dataset (Illumina Omni Express genotyping chip), and was well imputed across all other datasets (imputation quality score INFO ≥0.94, which is an estimate of genotype accuracy; the range of INFO scores is 0–1, where 1 indicates high confidence). All of these SNPs had minor

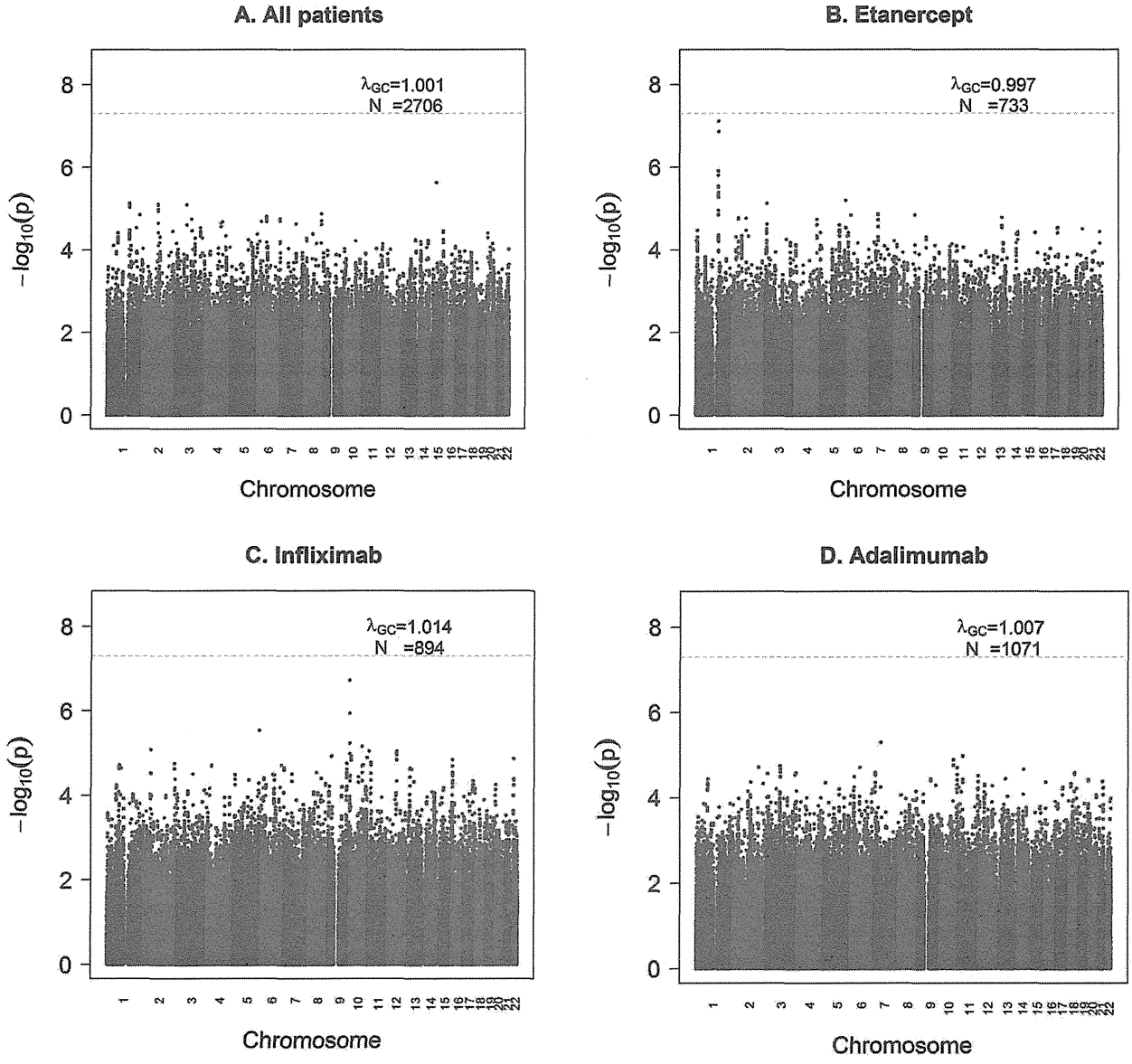
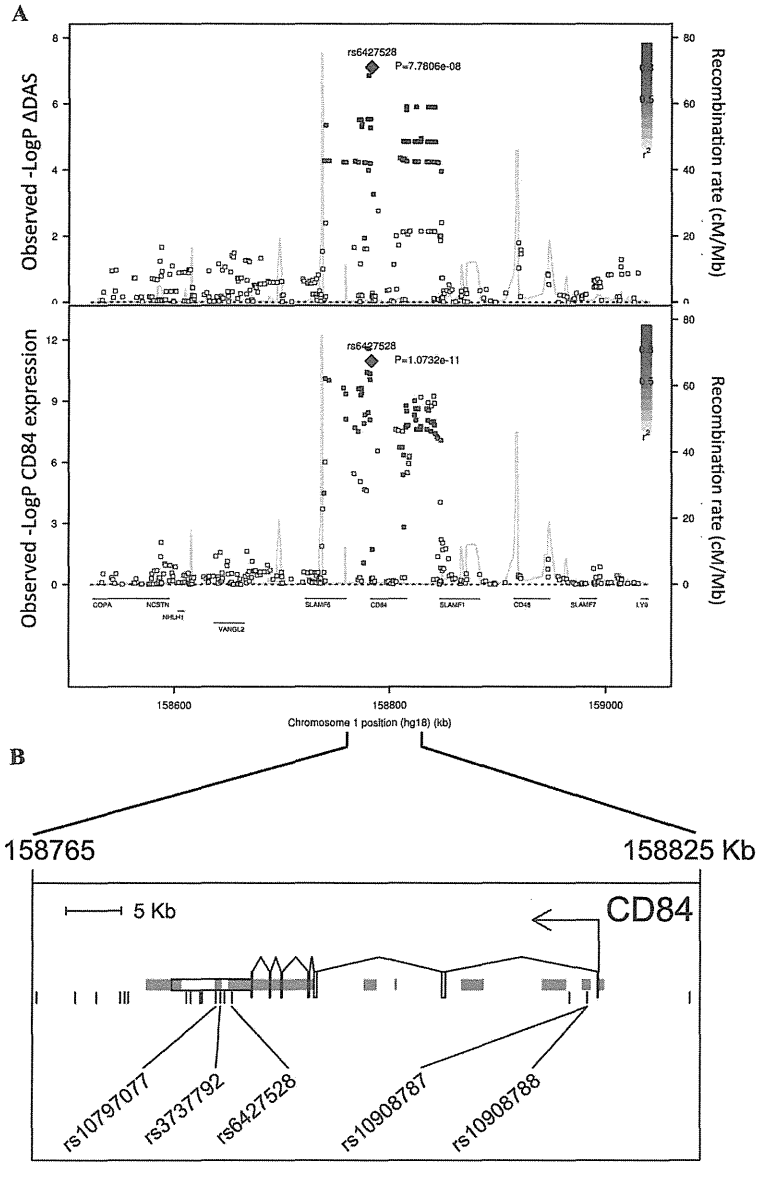


Figure 1. GWAS results for the ΔDAS phenotype. Shown are strengths of association (−Log10 P-value) for each SNP versus position along chromosomes 1 to 22. A) All samples (n = 2,706). B) Etanercept-treated patients (n = 733). C) Infliximab-treated patients (n = 894). D) Adalimumab-treated patients (n = 1,071).
doi:10.1371/journal.pgen.1003394.g001



C.

SNP (Major/ minor alleles)	Conserv-ation score[1]	DNase[2]	Transcription factor motifs altered		
			TF Motif	LOD(min) - LOD(maj)	Position weight matrix logo[3]
rs10797077 (G/A)	2.1	T-47D	AIRE_2	>6.6	
rs3737792 (G/A)	-1.3	-	-		
rs6427528 (A/G)	-2.8	-	KROX	>3	
			SREBP_4	-2.2	
rs10908787 (A/G)	-3.4	GM12878, Jurkat	-		
rs10908788 (A/G)	-7.8	GM12878	-		

Figure 2. Association results and SNP annotations in the *1q23 CD84* locus. A) Regional association plots with ΔDAS (top panel) and with *CD84* expression (bottom panel), showing strengths of association (−Log10 P-value) versus position (Kb) along chromosome 1. B) Schematic of *CD84* gene structure (RefSeq gene model, box exons connected by diagonal lines, arrow indicates direction of transcription) with strong enhancer chromatin states (orange rectangles) and SNPs in high LD (r2>0.8) with rs6427528 (vertical ticks). SNPs in enhancers are labeled below. C) Annotations of strong-enhancer rs6427528 proxy SNPs; listed are SNP rs-ID (major and minor alleles), conservation score, cell line with DNAse footprint if present, and transcription factor binding sites altered. 1- Genomic evolutionary rate profiling (GERP) conservation score, where a score >2 indicates conservation across mammals. 2- DNase footprint data are compiled from publicly available experiments by HaploReg. 3- Position weight matrix logos show transcription factor consensus binding sites with nucleotide bases proportional to binding importance. SNP position is boxed. Note that the rs10797077 AIRE_2 and the rs6427528 SREBP_4 motifs are on the minus strand (base complements correspond to SNP alleles), with the SREBP motif shown upside down to align with the rs6427528 KROX motif on the positive strand. Data are from HaploReg.
doi:10.1371/journal.pgen.1003394.g002

allele frequencies ranging from 7–10%. The SNP explains 2.6% variance in response to etanercept treatment.

For patients treated with infliximab, we observed a suggestive result on chromosome *10p14* (rs12570744, $P = 2\times10^{-7}$). No highly significant or suggestive results were observed for the ΔDAS phenotype in patients treated with adalimumab ($P_{META} > 10^{-5}$).

Qualitatively similar results were attained in the analysis of our secondary phenotype, EULAR good responder vs non-responder status (Figures S1 and S2). For SNPs at the *1q23* locus, the pattern of association with responder/non-responder status (etanercept-treated patients) was consistent with the results for ΔDAS ($P = 6\times10^{-3}$ for rs6427528 and rs1503860). We also identified potential novel associations, with suggestive results for infliximab (rs4336372, chromosome *5q35*, $P = 8\times10^{-7}$) and adalimumab (rs940928, chromosome *2q12*, $P = 2\times10^{-6}$).

eQTL and sequence analysis of the *CD84* gene

For each SNP with $P<10^{-6}$ identified by our GWAS (n = 6 independent SNPs), we searched for biological evidence to support a true positive association. We used genome-wide sequence data from the 1000 Genomes Project to search for putative functional variants in LD with the index SNP (defined as SNPs predicted to change protein-sequence or mRNA splicing). We also used genome-wide expression data to search for an expression quantitative trait locus (eQTL) in public databases and in peripheral blood mononuclear cells (PBMCs) in 228 non-RA patients and in 132 RA patients.

While we did not identify any variants disrupting protein-coding sequences or mRNA splicing, we did find that the *1q23* SNP associated with response to etanercept therapy was a strong eQTL in PBMCs (Figure 2A and Figure 3). In an analysis of 679 SNPs for cis-regulated expression of five genes in the region of LD (*SLAMF6*, *CD84*, *SLAMF1*, *CD48*, and *SLAMF7*), we found that rs6427528-*CD84* (and SNPs in LD with it) was the top eQTL of all results (n = 228 subjects; Figure 2A). This SNP was specifically associated with *CD84* expression, and was not an eQTL for other genes in the region (P>0.36 for the other genes).

We replicated our eQTL finding in 132 RA patients with both GWAS data and genome-wide expression data. PBMC expression data were available from RA patients in the Brigham RA Sequential Study (BRASS) and Autoimmune Biomarkers Collaborative Network (ABCoN) collections. We observed a significant association between rs6427528 genotype and *CD84* expression (linear regression adjusted for cohort P = 0.004, rank correlation P = 0.018). The direction of effect was the same as in the PBMC samples from 228 non-RA patients. A combined analysis of RA patients and the non-RA patient eQTL data (described above) yielded rank correlation $P = 3\times10^{-10}$ (n = 360 total individuals).

We searched sequence data to determine if rs6427528, or any of the SNPs in LD with it, were located within conserved, non-coding motifs that might explain the eQTL data. We used HaploReg [16] to examine the chromatin context of rs6427528 and 26 SNPs in LD with it (at $r^2>0.50$). We found that 5 SNPs occur in strong enhancers inferred from chromatin marks (Figure 2B) [17]. Two of these 5 SNPs, rs10797077 and rs6427528 ($r^2 = 0.74$ to each other), are predicted to disrupt transcription factor binding sites, and rs10797077 occurs at a site that shows conservation across mammalian genomes [18]. Figure 2C shows the DNA sequence position weight matrices of the transcription factor binding sites changed by rs10797077 (the minor allele creates a stronger binding site for the AIRE transcription factor) and rs6427528 (the minor allele creates a binding site for KROX and SREBP).

Expression of *CD84* as a biomarker of disease activity and treatment response

Because the genetic data demonstrates that the allele associated with better response is associated with higher *CD84* expression, this suggests that *CD84* expression itself may serve as a useful biomarker of disease activity or treatment response. We tested both hypotheses using PBMC expression data from the BRASS and ABCoN collections. First, we tested if *CD84* expression is associated with cross-sectional DAS, adjusting for age, gender and cohort (Figure 4). We observed a significant inverse association between *CD84* expression and cross-sectional DAS in 210 RA patients (beta = −0.3, P = 0.02, $r^2 = 0.02$). That is, higher *CD84* expression was associated with lower DAS, regardless of treatment.

Second, we tested *CD84* for association with our primary treatment response phenotype, ΔDAS. The sample size for this analysis was smaller than for the cross-sectional analysis, as we required that patients be on anti-TNF therapy and have pre- and post-treatment DAS. We found that *CD84* expression levels showed a non-significant trend towards an association with ΔDAS in 31 etanercept-treated patients (beta = 0.2, $r^2 = 0.002$, P = 0.46) and in all 78 anti-TNF-treated patients (beta = 0.14, $r^2 = 0.004$, P = 0.4). The effect is in the same direction one would predict based on the genetic association at rs6427528: the allele associated with better response is also associated with higher *CD84* expression (Figure 3), and in 31 RA patients, higher *CD84* expression (regardless of genotype) is associated with a larger ΔDAS (i.e., better response; Figure 4).

Replication of genetic data in a small, multi-ethnic cohort

Since most of the samples available to us as part of our international collaboration were included in our GWAS, few additional samples were available for replication. In addition, the remaining samples available to us were from different ethnic backgrounds. Nonetheless, we sought to replicate the associations of rs6427528 with ΔDAS in these additional samples. We genotyped 139 etanercept-treated patients from a rheumatoid arthritis registry in Portugal (Reuma.pt) and 151 etanercept-treated patients from two Japanese collections (IORRA, n = 88 patients on etanercept and Kyoto University, n = 63 on etanercept). Replication sample sizes, clinical data and results for these

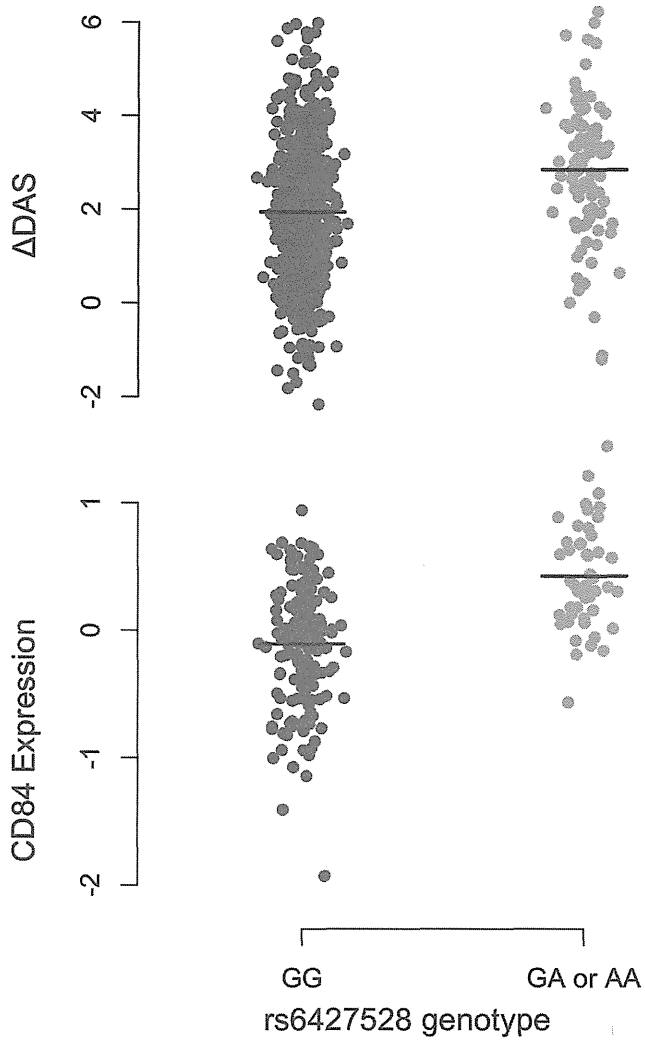


Figure 3. *1q23/CD84* genotype association plots for ΔDAS and *CD84* gene expression. Shown are ΔDAS in our GWAS in etanercept-treated patients (top panel, n = 733; n = 634 with the GG genotype and n = 99 with the GA or AA genotype) and *CD84* expression in our eQTL results (bottom panel, n = 228 non-RA patients; n = 178 with the GG genotype and n = 50 with the GA or AA genotype). The rare-allele homozygous genotype AA was observed four times in our ΔDAS GWAS and was pooled with the heterozygous GA genotype for this figure; AA homozygotes were not observed in the *CD84* eQTL data. Association analyses reported in the text regressed phenotype (ΔDAS, $P = 8 \times 10^{-8}$; *CD84* expression, $P = 1 \times 10^{-11}$) on minor-allele dosage (range 0–2).
doi:10.1371/journal.pgen.1003394.g003

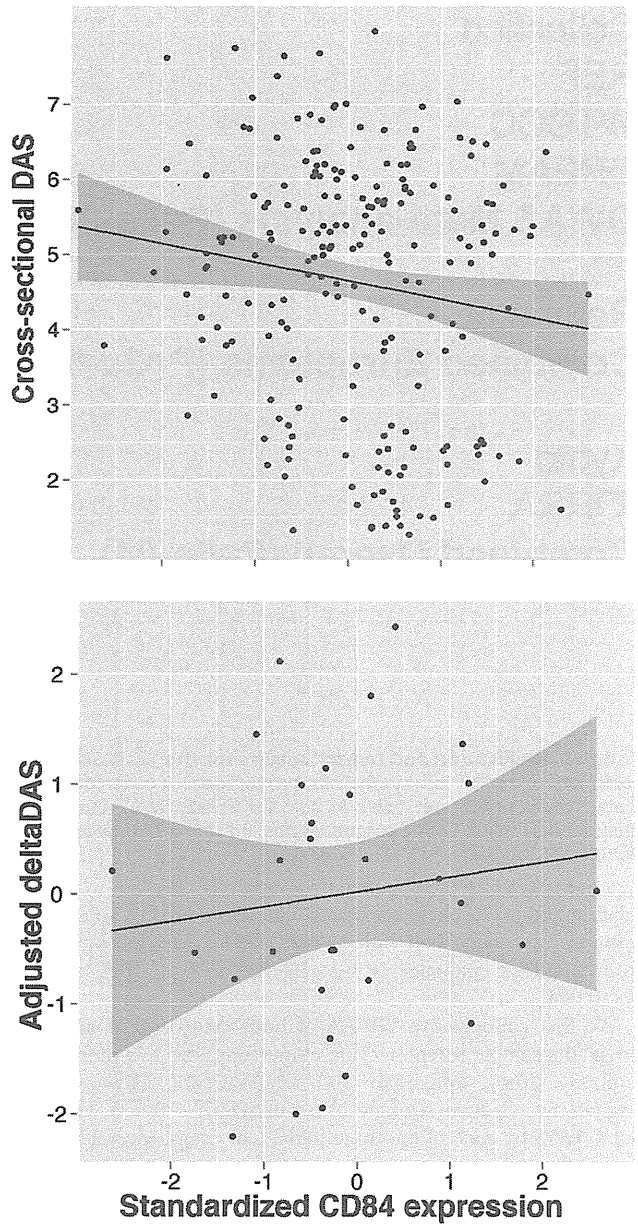


Figure 4. *CD84* expression level and clinical features. Analyses are shown in RA patients from the BRASS and ABCoN registries, for baseline DAS (top panel, n = 210; $R^2 = 0.02$, p = 0.02) and ΔDAS (bottom panel, n = 31; $R^2 = 0.001$, p = 0.46). Best-fit linear regression lines are shown in black, with shaded regions showing linear regression model (slope and intercept) 95% confidence intervals. *CD84* expression levels were quantile normalized, and ΔDAS values were adjusted for age, gender and baseline DAS.
doi:10.1371/journal.pgen.1003394.g004

two SNPs are shown in Table S4. Based on the observed effect size in the GWAS and observed allele frequency in the replication samples, we had 32% power to replicate this finding in the Portuguese samples and 17% power to replicate this finding in the Asian samples at $P<0.05$. The same association analysis as for GWAS was carried out: linear regression assuming an additive genetic model and using ΔDAS as phenotype, adjusted for baseline DAS. Replication results are shown in Figure 5.

While the SNPs fail to replicate in these patient collections at $P<0.05$, the direction of effect is the same in the Portuguese and Kyoto replication samples as in our GWAS. In a combined analysis limited to subjects of European-ancestry (GWAS data and Portuguese replication samples), rs6427528 remained highly suggestive ($P = 2 \times 10^{-6}$). Including the Japanese subjects, the overall GWAS+replication combined meta-analysis P-value remained suggestive ($P = 5 \times 10^{-4}$).

Discussion

Here we present the largest GWAS to date on anti-TNF therapy response in 2,706 RA patients. We find a significant association at the *1q23/CD84* locus in 733 etanercept treated patients ($P = 8 \times 10^{-8}$), but not in RA patients treated with drugs that act as a monoclonal antibody to neutralize TNF (infliximab or

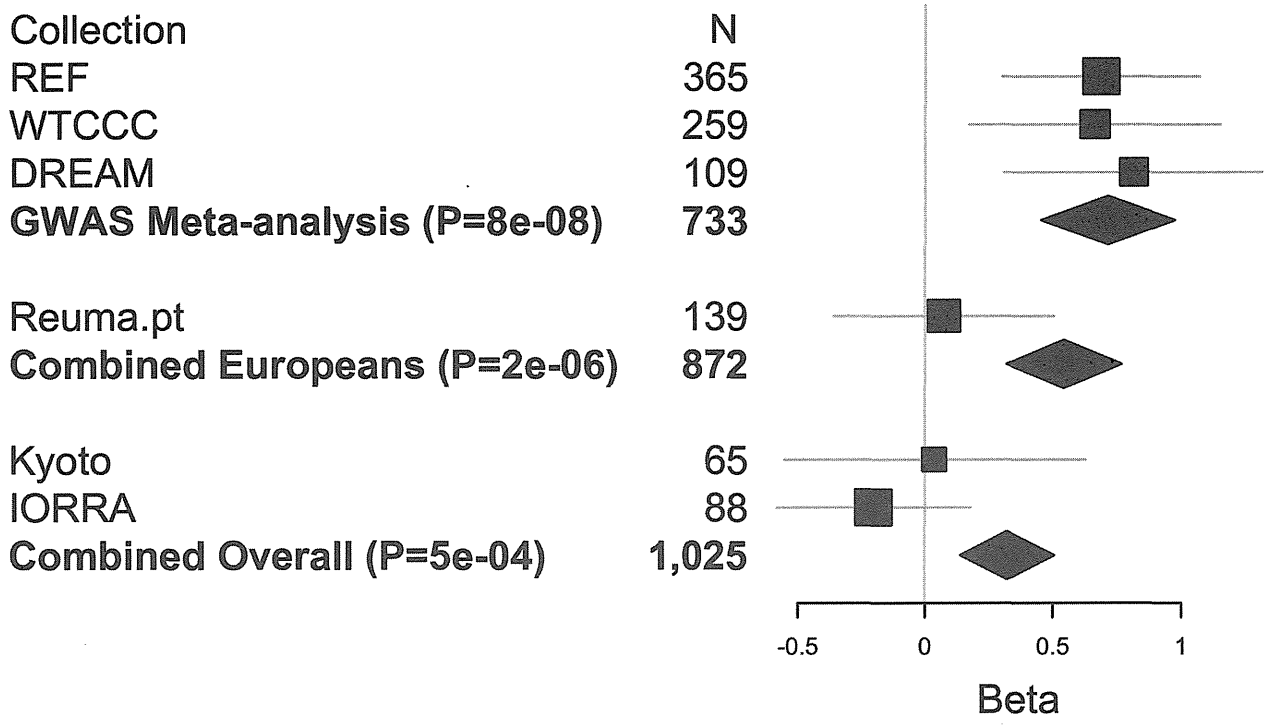


Figure 5. Replication and overall results for the *CD84* SNP rs6427528. Forest plot shows each cohort, sample size and linear regression beta coefficient estimates with symbol size proportional to cohort sample size and thin horizontal lines showing beta 95% CIs. Inverse variance weighted meta-analysis results are shown in bold for GWAS, GWAS+European (Portuguese) replication samples, and for GWAS+European+Asian (Japanese) replication samples, with diamond widths indicating beta 95% CIs.
doi:10.1371/journal.pgen.1003394.g005

adalimumab). The allele associated with a larger ΔDAS (i.e., better response) was associated with higher *CD84* expression in PBMCs from non-RA patients ($P = 1 \times 10^{-11}$) and in RA patients ($P = 0.004$).

We first conducted a GWAS of both categories of anti-TNF drugs (the soluble receptor drug, etanercept, and two monoclonal antibody drugs, infliximab and adalimumab). However, this analysis revealed no strongly associated SNPs. When we subset our GWAS by each of the three individual drugs, several SNPs in the *1q23* locus were highly significant in etanercept-treated patients, and SNPs in three other loci (*10p15*, *5q35* and *2q12*) were associated in infliximab or adalimumab subset analyses. Furthermore, the top SNPs for each analysis (Table S3) showed little correlation across the three anti-TNF drugs. This simple observation suggests that genetic control of treatment response may be different for different drugs. This finding is consistent with the clinical observation that RA patients who fail one anti-TNF drug may still respond to a different anti-TNF drug, albeit at lower rates of response [19]. If confirmed in larger samples and more comprehensive analyses, then this could have major implications for how physicians prescribe these drugs.

The most significant finding from our GWAS was a set of equivalent SNPs in LD with each other from the *1q23* locus in etanercept-treated RA patients (Figure 1 and Figure 2A). While the top SNP did not reach genome-wide significance in predicting treatment response, it did reach genome-wide significance as an eQTL in PBMCs ($P = 1 \times 10^{-11}$; Figure 2A). This finding indicates that the SNP (or another variant in LD with it) is indeed biologically functional in a human tissue that is important in the immune response. Two SNPs, rs10797077 and rs6427528, disrupt transcription factor binding sites, and represent excellent candidates for the causative allele to explain the effect on *CD84* expression (Figure 2C).

Our findings suggest that *CD84* genotype and/or expression could be a biomarker for etanercept treatment response among individuals of European ancestry. The genetic and expression data predict that *CD84* expression should be positively associated with treatment response (i.e., higher expression is associated with better response; Figure 3). While we did not observe a significant association between *CD84* levels and ΔDAS, we did observe a trend consistent with this prediction (Figure 4). Importantly, we note that power was extremely limited with the small sample sizes for which we had *CD84* expression as well as drug response data ($n = 31$ RA patients treated with etanercept).

The *CD84* gene is a compelling candidate for immune response, belonging to the CD2 subset of the immunoglobulin superfamily. It has been implicated in T-cell activation and maturation [20]. CD84 localizes to the surface of CD4+ and CD8+ T cells, and acts as a costimulatory molecule for IFN-gamma secretion [21]. *CD84* is also expressed in B-cells, monocytes and platelets. *CD84* has not been previously implicated in genetic studies of RA risk, disease activity, disease severity, or treatment response.

A limitation of our study is the small sample size available for replication ($n = 290$ etanercept-treated patients), and the lack of replication observed for the top *CD84* SNP (rs6427528) among patients of Portuguese and Japanese ancestry. The simplest explanation is that our original observation in the GWAS data represents a false positive association. However, the eQTL and gene expression data argue against this possibility. Explanations for a false negative finding in our replication collections include: (1) lack of power, especially if the effect size observed in the GWAS represents an over-estimate of the true effect size (the Winner's Curse) – we estimate that we had 32% and 17% power (at $P = 0.05$) to detect an association in the Portuguese and Japanese sample collections, respectively; (2) clinical heterogeneity, which is