METHODS

Subjects. Our study included 29,880 RA cases (88.1% seropositive and 9.3% seronegative for anti-citrullinated peptide antibody (ACPA) or rheumatoid factor (RF), and 2.6% who had unknown autoantibody status) and 73,758 controls. All RA cases fulfilled the 1987 criteria of the American College of Rheumatology for RA diagnosis[24], or were diagnosed with RA by a professional rheumatologist. The 19,234 RA cases and 61,565 controls enrolled in the stage 1 trans-ethnic GWAS meta-analysis were obtained from 22 studies on people with European and Asian ancestries (14,361 RA cases and 43,923 controls from 18 studies of Europeans and 4,873 RA cases and 17,642 controls from 4 studies of Asians): BRASS[2], CANADA[2], EIRA[2], NARAC1[2], NARAC2[2], WTCCC[2], Rheumatoid Arthritis Consortium International for Immunochip (RACI)-UK[4], RACI-US[4], RACI-SE-E[4], RACI-SE-U[4], RACI-NL[4], RACI-ES[4], RACI-i2b2, ReAct, Dutch (including AMC, BeSt, LUMC and DREAM), anti-TNF response to therapy collection (ACR-REF: BRAGGSS, BRAGGSS2, ERA, KI and TEAR), CORRONA, Vanderbilt, three studies from the GARNET consortium (BioBank Japan Project[3], Kyoto University[3] and IORRA[3]), and Korea. Of these, GWAS data of 4,309 RA cases and 8,700 controls from six studies (RACI-i2b2, ReAct, Dutch, ACR-REF, CORRONA and Vanderbilt) have not been previously published.

The 3,708 RA cases and 5,535 controls enrolled in the stage 2 *in silico* replication study were obtained from two studies of Europeans (2,780 RA cases and 4,700 controls from Genentech and SLEGEN) and Asians (928 RA cases and 835 controls from China) (H.X. *et al.*, manuscript submitted). The 6,938 RA cases and 6,658 controls enrolled in the stage 3 *de novo* replication study were obtained from two studies of Europeans (995 RA cases and 1,101 controls from CANADAII[2]) and Asians (5,943 RA cases and 5,557 controls from BioBank Japan Project, Kyoto University and IORRA[3]).

All subjects in the stage 1, stage 2 and stage 3 studies were confirmed to be independent through analysis of overlapping SNP markers. Any duplicate subjects were removed from the stage 2 and stage 3 replication studies, leading to slightly different sample sizes compared with previous studies that used these same collections[2,3].

All participants provided written informed consent for participation in the study as approved by the ethical committees of each of the institutional review boards. Detailed descriptions of the study design, participating cohorts and the clinical characteristics of the RA cases are provided in detail in Extended Data Fig. 1 and Extended Data Table 1a, as well as in previous reports[2–4].

Genotyping, quality control and genotype imputation of GWAS data. Genotyping platforms and quality control criteria of GWAS, including cut-off values for sample call rate, SNP call rate, minor allele frequency (MAF), and Hardy–Weinberg equilibrium (HWE) *P* value, covariates in the analysis, and imputation reference panel information are provided for each study in Extended Data Table 1b. All studies were analysed based on the same analytical protocol, including exclusion of closely related subjects and outliers in terms of ancestries, as described elsewhere[3]. After applying quality control criteria, whole-genome genotype imputation was performed using 1000 Genomes Project Phase I (α) European ($n = 381$) and Asian ($n = 286$) data as references[11]. We excluded monomorphic or singleton SNPs or SNPs with deviation of HWE ($P < 1.0 \times 10^{-7}$) from each of the reference panels. GWAS data were split into ~300 chunks that evenly covered whole-genome regions and additionally included 300 kb of duplicated regions between neighbouring chunks. Immunochip data were split into ~2,000 chunks that included each of the targeted regions or SNPs on the array. Each chunk was pre-phased and imputed by using minimac (release stamp 2011-10-27). SNPs in the X chromosome were imputed for males and females separately. We excluded imputed SNPs that were duplicated between chunks, SNPs with MAF < 0.005 in RA cases or controls, or with low imputation score (Rsq < 0.5 for genome-wide array and < 0.7 for Immunochip) from each study. We found that imputation of Immunochip effectively increased the number of the available SNPs by 7.0 fold (from ~129,000 SNPs to ~924,000 SNPs) to cover ~12% of common SNPs (MAF > 0.05) included in the 1000 Genomes Project reference panel for European ancestry[11].

Stage 1 trans-ethnic genome-wide meta-analysis. Associations of SNPs with RA were evaluated by logistic regression models assuming additive effects of the allele dosages including top 5 or 10 principal components as covariates (if available) using mach2dat v.1.0.16 (Extended Data Table 1b). Allele dosages of the SNPs in X chromosome were assigned as 0/1/2 for females and 0/2 for males and analysed separately. Meta-analysis was performed for the trans-ethnic study (both Europeans and Asians), European study, and Asian study separately. The SNPs available in ≥3 studies were evaluated in each GWAS meta-analysis, which yielded ~10 million autosomal and X-chromosomal SNPs. Information about the SNPs, including the coded alleles, was oriented to the forward strand of the NCBI build 37 reference sequence. Meta-analysis was conducted by an inverse-variance method assuming a fixed-effects model on the effect estimates (β) and the standard errors of the allele dosages using the Java source code implemented by the authors[25]. Double GC correction was carried out using the inflation factor (λ_{GC}) obtained from the results of each GWAS and the GWAS meta-analysis[25] after removing the SNPs located ± 1 Mb from known RA loci or in the MHC region (chromosome 6, 25–35 Mb). Although there is not yet uniform consensus on the application of double GC correction, we note that potential effects of double GC correction would not be substantial in our study because of the small values of the inflation factors in the GWAS meta-analysis ($\lambda_{GC} < 1.075$ and λ_{GC} adjusted for 1,000 cases and 1,000 controls ($\lambda_{GC\_1,000}$) < 1.005; Extended Data Table 1b).

As for the definition of known RA risk loci in this study, we included the loci that showed significant associations in one of the previous studies ($P < 5.0 \times 10^{-8}$) or that had been replicated in independent cohorts. We consider the locus including multiple independent signals of associations as a single locus, such as the MHC locus[12] and *TNFAIP3* (ref. 4). Although 6 of these 59 loci previously identified as known RA risk loci did not reach a suggestive level of association (defined as $P < 5.0 \times 10^{-6}$) in our stage 1 meta-analysis, previous studies have gone on to replicate most of these associations in additional samples (Supplementary Table 1)[2,3]. Thus, the number of confirmed RA risk loci is 101 (including the MHC region).

Stage 2 and stage 3 replication studies. *In silico* (stage 2) and *de novo* (stage 3) replication studies were conducted using independent European and Asian subjects (Extended Data Table 1). The 146 loci that satisfied $P < 5.0 \times 10^{-6}$ in the stage 1 trans-ethnic, European or Asian GWAS meta-analysis were selected for the stage 2 *in silico* replication study. The SNPs that demonstrated the most significant associations were selected from each of the loci. When the SNP was not available in replication data sets, a proxy SNP with the highest linkage disequilibrium ($r^2 > 0.80$) was alternatively assessed. GWAS quality control, genotype imputation and association analysis were assessed in the same manner as in the stage 1 GWAS. For the 60 loci that demonstrated suggestive associations in the combined results of the stage 1 GWAS meta-analysis and the stage 2 *in silico* replication study but were not included as a known RA risk locus, we calculated statistical power to newly achieve a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ for Europeans and Asians separately, which were estimated based on the allele frequencies, ORs and *de novo* replication sample sizes of the populations. We then selected the top 20 SNPs with the highest statistical power for Europeans and Asians separately (in total 32 SNPs), and conducted the stage 3 *de novo* replication study. Genotyping methods, quality control and confirmation of subject independence in the stage 3 *de novo* replication study were described previously[2,3]. The combined study of the stage 1 GWAS meta-analysis and the stages 2 and 3 replication studies was conducted by an inverse-variance method assuming a fixed-effects model[25].

Trans-ethnic and functional annotations of RA risk SNPs. Trans-ethnic comparisons of RAF (in the reference panels), ORs and explained heritability were conducted using the results of the stage 1 GWAS meta-analysis of Europeans and Asians. Correlations of RAF and OR were evaluated using Spearman's correlation test. ORs were defined based on minor alleles in Europeans. Explained heritability was estimated by applying a liability-threshold model assuming disease prevalence of 0.5% (ref. 10) and using the RAF and OR of the population(s) according to the genetic risk model. For the population-specific genetic risk model, the RAF and OR of the same population was used. For the trans-ethnic genetic risk model, the RAF of the population but the OR of the other population was used.

Details of the overlap enrichment analysis of RA risk SNPs with H3K4me3 peaks have been described elsewhere[13]. Briefly, we evaluated whether the RA risk SNPs (outside of the MHC region) and SNPs in linkage disequilibrium ($r^2 > 0.80$) with them were enriched in overlap with H3K4me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) assay peaks of 34 cell types obtained from the National Institutes of Health Roadmap Epigenomics Mapping Consortium, by a permutation procedure with $\times 10^5$ iterations.

Fine mapping of causal risk alleles. For fine mapping of the causal risk alleles, we selected the 31 RA risk loci where the risk SNPs yielded $P < 1.0 \times 10^{-3}$ in the stage 1 GWAS meta-analysis of both Europeans and Asians with the same directional effects of alleles (outside of the MHC region). For fine mapping using linkage-disequilibrium structure differences between the populations, we calculated average numbers of the SNPs in linkage disequilbrium ($r^2 > 0.80$) in Europeans, Asians, and in both Europeans and Asians, separately.

For fine mapping using H3K4me3 peaks of T_{reg} primary cells, we first evaluated H3K4me3 peak overlap enrichment of the SNPs in linkage disequilbrium (in Europeans and Asians) compared with the neighbouring SNPs (±2 Mb). We fixed the SNP positions but physically slid H3K4me3 peak positions by 1 kb bins within ±2 Mb regions of the risk SNPs, and calculated overlap of the SNPs in linkage disequilibrium with H3K4me3 peaks for each sliding step, and evaluated the significance of overlap in the original peak positions by a one-sided exact test assuming enrichment of overlap. For the 10 loci that demonstrated significant overlap ($P < 0.05$), we calculated the average number of the SNPs that were in linkage disequilibrium in both Europeans and Asians and also included in H3K4me3 peaks.

Pleiotropy analysis. We downloaded phenotype-associated SNPs and phenotype information from the National Human Genome Research Institute (NHGRI) GWAS catalogue database[26] on 31 January, 2013. We selected 4,676 significantly associated SNPs ($P < 5.0 \times 10^{-8}$) corresponding to 311 phenotypes (other than RA). We manually curated the phenotypes by combining the same but differently named phenotypes into a single phenotype (for example, from 'urate levels', 'uric acid levels' and 'renal function-related traits (urea)' to 'urate levels'), or splitting merged phenotypes into sub-categorical phenotypes (for example, from 'white blood cell types' into 'neutrophil counts', 'lymphocyte counts', 'monocyte counts', 'eosinophil counts' or 'basophil counts'). Lists of curated phenotypes and SNPs are available at http://plaza.umin.ac.jp/~yokada/datasource/software.htm.

For each of the selected NHGRI GWAS catalogue SNPs and the RA risk SNPs identified by our study (located outside of the MHC region), we defined the genetic region based on ±25 kb of the SNP or the neighbouring SNP positions in moderate linkage disequilibrium with it in Europeans or Asians ($r^2 > 0.50$). If multiple different SNPs with overlapping regions were registered for the same phenotype, they were merged into a single region. We defined 'region-based pleiotropy' as two phenotype-associated SNPs sharing part of their genetic regions or sharing any UCSC hg19 reference gene(s) that partly overlapped each of the regions (Extended Data Fig. 4a). We defined 'allele-based pleiotropy' as two phenotype-associated SNPs that were in linkage disequilibrium in Europeans or Asians ($r^2 > 0.80$). We defined the direction of an effect as 'concordant' with RA risk if the RA risk allele also leads to increased risk of the disease or increased dosage of the quantitative trait; similarly, we defined relationships as 'discordant' if the RA risk allele is associated with decreased risk of the disease phenotype (or if the RA risk allele leads to decreased dosage of the quantitative trait).

We evaluated statistical significance of region-based pleiotropy of the registered phenotypes with RA by a permutation procedure with $\times 10^7$ iterations. When one phenotype had n loci of which m loci were in region-based pleiotropy with RA, we obtained a null distribution of m by randomly selecting n SNPs from obtained NHGRI GWAS catalogue data and calculating the number of the observed region-based pleiotropy with RA for each of the iteration steps. For estimation of the null distribution, we did not include the SNPs associated with several autoimmune diseases that were previously reported to share pleiotropic associations with RA (Crohn's disease, type 1 diabetes, multiple sclerosis, coeliac disease, systemic lupus erythematosus, ulcerative colitis and psoriasis)[2].

Prioritization of biological candidate genes from RA risk loci. For RA risk SNPs outside of the MHC region, functional annotations were conducted by Annovar (hg19). RA risk SNPs were classified if any of the SNPs in linkage disequilibrium ($r^2 > 0.80$) in Europeans or Asians were annotated in order of priority of missense (or nonsense), synonymous or non-coding (with or without *cis*-eQTL) SNPs. We also applied this SNP annotation scheme to 10,000 randomly selected genome-wide common SNPs (MAF > 0.05 in Europeans or Asians).

We then assessed *cis*-eQTL effects by referring two eQTL data sets: the study for peripheral blood mononuclear cells (PBMCs) obtained from 5,311 European subjects[6] and newly generated cell-specific eQTL analysis for $CD4^+$ T cells and $CD14^+CD16^-$ monocytes from 212 European subjects (ImmVar project; T.R. *et al.*, manuscript submitted). When the RA risk SNP was not available in eQTL data sets, we alternatively used the results of best proxy SNPs in linkage disequilibrium with the highest r^2 value (>0.80). We applied the significance thresholds defined in the original studies (FDR $q < 0.05$ for PBMC eQTL and gene-based permutation $P < 0.05$ for cell-specific eQTL).

We obtained PID genes and their classification categories as defined by the International Union of Immunological Societies Expert Committee[14], downloaded cancer somatic mutation genes from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[15], and downloaded knockout mouse phenotype labels and gene information from the Mouse Genome Informatics (MGI) database[16] on 31 January, 2013 (Supplementary Tables 2–5). We defined 377 RA risk genes included in the 100 RA risk loci (outside of the MHC region) according to the criteria described in the previous section (±25 kb or $r^2 > 0.50$), and evaluated overlap with PID categories, cancer phenotypes with registered somatic mutations, and phenotype labels of knockout mouse genes with human orthologues. Statistical significance of enrichment in gene overlap was assessed by a permutation procedure with $\times 10^6$ iterations. For each iteration step, we randomly selected 100 genetic loci matched for number of nearby genes with those in non-MHC 100 RA risk loci. When one gene category had m genes overlapping with RA risk genes, we obtained a null distribution of m by calculating the number of genes in the selected loci overlapping with RA risk genes for each iteration step.

We conducted molecular pathway enrichment analysis using MAGENTA software[9] and adopting Ingenuity and BIOCARTA databases as pathway information resources. We conducted two patterns of analyses by inputting genome-wide SNP P values of the current trans-ethnic meta-analysis (stage 1) and the previous meta-analysis of RA[2] separately. As the previous meta-analysis was conducted using imputed data based on HapMap Phase II panels, we re-performed the meta-analysis using the same subjects but with newly imputed genotype data based on the 1000 Genomes Project reference panel[11] to make SNP coverage conditions identical between the meta-analyses. Significance of the molecular pathway was evaluated by FDR q values obtained from $\times 10^5$ iterations of permutations.

We scored each of the genes included in the RA risk loci (outside of the MHC region) by adopting the following eight selection criteria and calculating the number of the satisfied criteria: (1) genes for which RA risk SNPs or any of the SNPs in linkage disequilibrium ($r^2 > 0.80$) with them were annotated as missense variants; (2) genes for which significant *cis*-eQTL of any of PBMCs, T cells or monocytes were observed for RA risk SNPs (FDR $q < 0.05$ for PBMCs and permutation $P < 0.05$ for T cells and monocytes); (3) genes prioritized by PubMed text mining using GRAIL[7] with gene-based $P < 0.05$; (4) genes prioritized by PPI network using DAPPLE[8] with gene-based $P < 0.05$; (5) PID genes[14]; (6) haematological cancer somatic mutation genes[15]; (7) genes for which ≥2 of associated phenotype labels ('haematopoietic system phenotype', 'immune system phenotype' and 'cellular phenotype'; $P < 1.0 \times 10^{-4}$) were observed for knockout mouse[16]; and (8) genes prioritized by molecular pathway analysis using MAGENTA[9], which were included in the significantly enriched pathways (FDR $q < 0.05$) with gene-based $P < 0.05$. Because these criteria showed weak correlations with each other ($R^2 < 0.26$; Extended Data Fig. 6c), each gene was given a score based on the number of criteria that were met (scores ranging from 0–8 for each gene). We defined the genes with a score ≥2 as 'biological RA risk genes'.

For each gene in RA risk loci, we evaluated whether the gene was the nearest gene to the RA risk SNP within the risk locus, or whether the RA risk SNP (or SNPs in linkage disequilibrium with it) of the gene overlapped with H3K4me3 histone peaks of cell types. The difference in proportions of genes that were the nearest gene to biological RA risk genes (score ≥2) and non-biological genes (score <2) was evaluated by using Fisher's exact test implemented in R statistical software (v.2.15.2). The difference in the proportions of genes overlapping with T_{reg} primary cell H3K4me3 peaks between biological and non-biological genes was assessed by a permutation procedure by shuffling the overlapping status of RA risk SNPs/loci with $\times 10^5$ iterations.

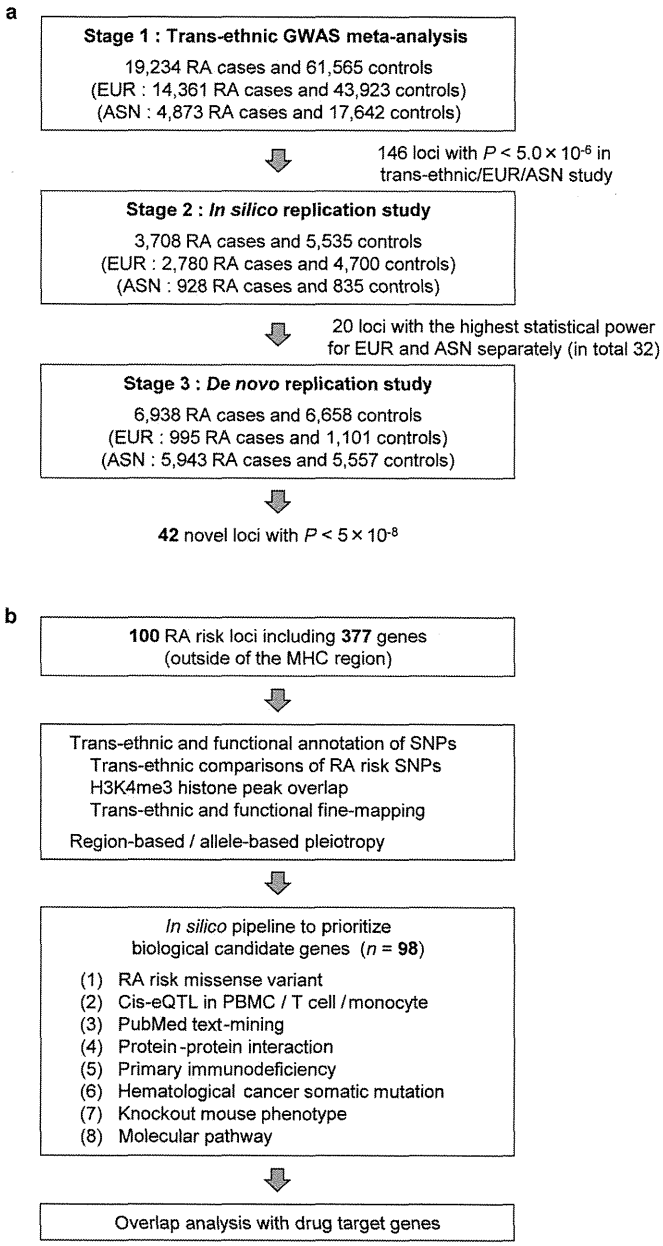
Drug target gene enrichment analysis. We obtained drug target genes and corresponding drug information from DrugBank[17] and the Therapeutic Targets Database (TTD)[18] on 31 January, 2013, as well as additional literature searches. We selected drug target genes that had pharmacological activities (for the genes from DrugBank) and human orthologues, and that were annotated to any of the approved, clinical trial or experimental drugs (Supplementary Table 6). We manually extracted drug target genes annotated to approved RA drugs on the basis of discussions with professional rheumatologists (Extended Data Fig. 7a). We extracted genes in direct PPI with biological RA risk genes by using the InWeb database[27]. To take account of potential dependence between PPI genes and drug target genes, overlap of biological RA risk genes and genes in direct PPI with them with drug target genes was assessed by a permutation procedure with $\times 10^5$ iterations.

Let x be the set of the biological RA risk genes and genes in direct PPI with them (n_x genes), y be the set of genes with protein products that are the direct target of approved RA drugs (n_y genes), and z be the set of genes with protein products that are the direct target of all approved drugs (n_z genes). We defined $n_{x\cap y}$ and $n_{x\cap z}$ as the numbers of genes overlapping between x and y and between x and z, respectively. For each of 10,000 iteration steps, we randomly selected a gene set of x' including n_x genes from the entire PPI network (12,735 genes). We defined $n_{x\cap y}'$ and $n_{x\cap z}'$ as the numbers of genes overlapping between x' and y, and between x' and z, respectively. The distributions of $n_{x\cap y}'$, $n_{x\cap z}'$ and $n_{x\cap y}'/n_{x\cap z}'$ obtained from the total iterations were defined as the null distributions of $n_{x\cap y}$, $n_{x\cap z}$, and $n_{x\cap y}/n_{x\cap z}$, respectively. Fold enrichment of overlap with approved RA drug target genes was defined as $n_{x\cap y}/m(n_{x\cap y}')$, where $m(t)$ represents the mean value of the distribution of t. Fold enrichment of overlap with approved all drug target genes was defined as $n_{x\cap z}/m(n_{x\cap z}')$. Relative fold enrichment of overlap with RA drug target genes and with all drug target genes was defined as $(n_{x\cap y}/n_{x\cap z})/m(n_{x\cap y}'/n_{x\cap z}')$. Significance of the enrichment was evaluated by one-sided permutation tests examining $n_{x\cap y}$, $n_{x\cap z}$, and $n_{x\cap y}/n_{x\cap z}$ in their null distributions.

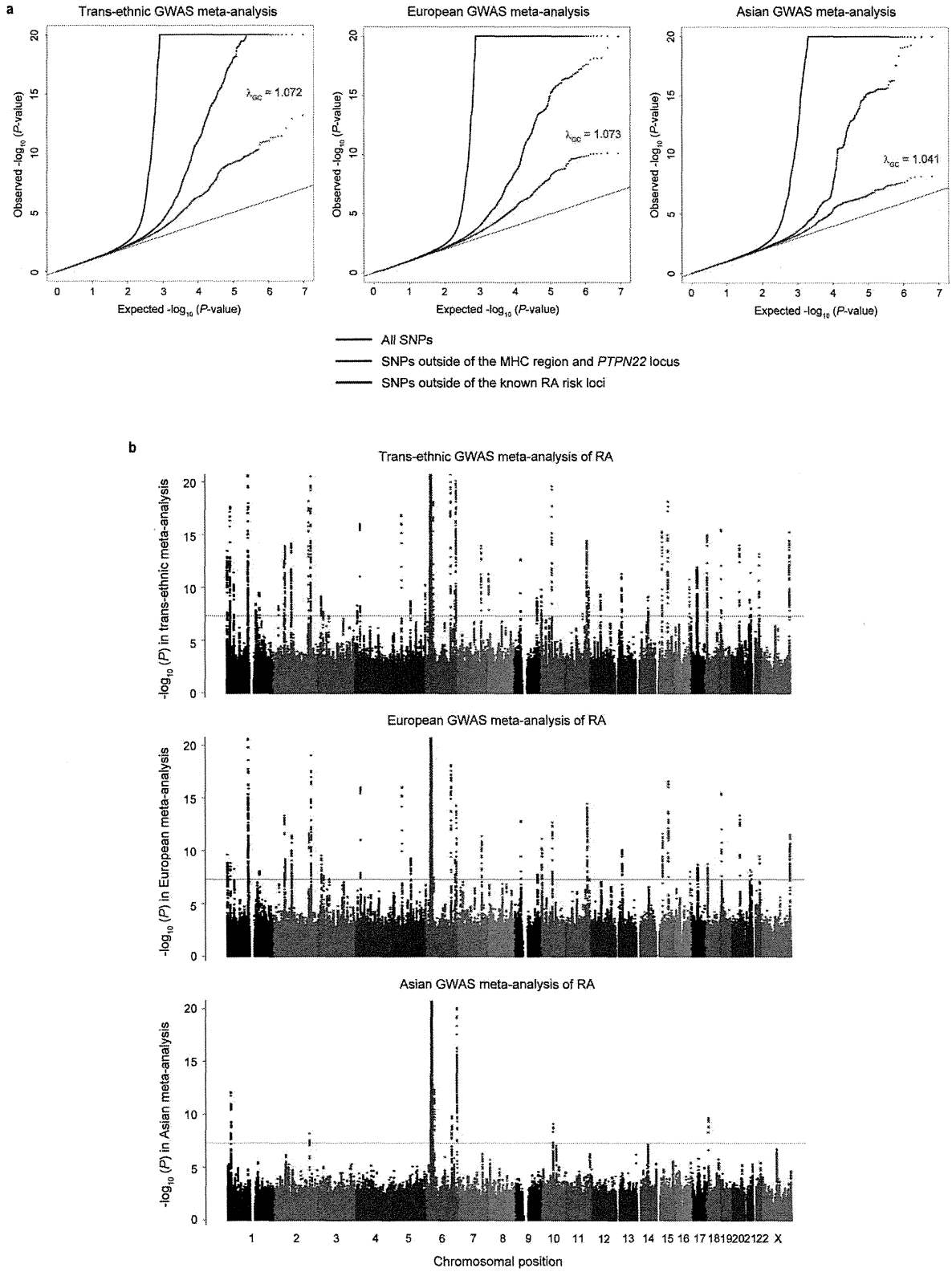
Web resources. The following websites provide valuable additional resources. Summary statistics from the GWAS meta-analysis, source codes, and data sources have been deposited at http://plaza.umin.ac.jp/~yokada/datasource/software.htm; GARNET consortium, http://www.twmu.ac.jp/IOR/garnet/home.html; i2b2, https://www.i2b2.org/index.html; SLEGEN, http://www.lupusresearch.org/lupus-research/slegen.html; 1000 Genomes Project, http://www.1000genomes.org/; minimac, http://genome.sph.umich.edu/wiki/Minimac; mach2dat, http://www.sph.umich.edu/csg/abecasis/MACH/index.html; Annovar, http://www.openbioinformatics.org/annovar/; ImmVar, http://www.immvar.org/; NIH Roadmap Epigenomics Mapping Consortium, http://www.roadmapepigenomics.org/; NHGRI GWAS catalogue, http://www.genome.gov/GWAStudies/; COSMIC, http://cancer.sanger.ac.uk/cancergenome/projects/

cosmic/; MGI, http://www.informatics.jax.org/; MAGENTA, http://www.broadinstitute.org/mpg/magenta/; Ingenuity, http://www.ingenuity.com/; BIOCARTA, http://www.biocarta.com/; GRAIL, http://www.broadinstitute.org/mpg/grail/; DAPPLE, http://www.broadinstitute.org/mpg/dapple/dapple.php; R statistical software, http://www.r-project.org/; DrugBank, http://www.drugbank.ca/; TTD, http://bidd.nus.edu.sg/group/ttd/ttd.asp.

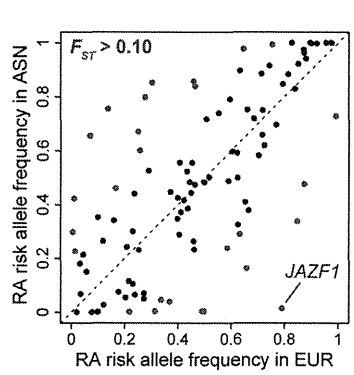
24. Arnett, F. C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31,** 315–324 (1988).
25. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nature Genet.* **44,** 904–909 (2012).
26. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
27. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnol.* **25,** 309–316 (2007).
28. Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423,** 506–511 (2003).
29. Elliott, P. *et al.* Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *J. Am. Med. Assoc.* **302,** 37–48 (2009).
30. Cortes, A. *et al.* Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nature Genet.* **45,** 730–738 (2013).

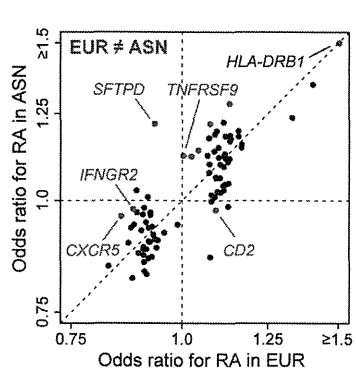


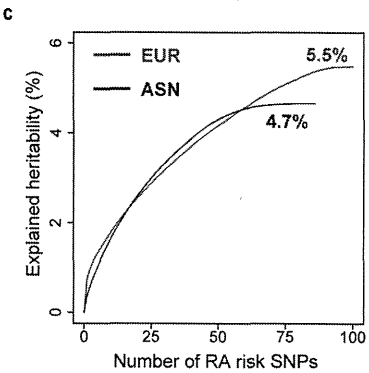
Extended Data Figure 1 | An overview of the study design. a, We conducted a three-stage trans-ethnic meta-analysis in total of 29,880 RA cases and 73,758 controls of European (EUR) and Asian (ASN) ancestry. The stage 1 GWAS meta-analysis included 19,234 RA cases and 61,565 controls from 22 studies, which was followed by the stage 2 *in silico* replication study (3,708 RA cases and 5,535 controls) and stage 3 *de novo* replication study (6,938 RA cases and 6,658 controls). In the combined study of stages 1–3, we identified 42 novel RA risk loci, which increased the total number of RA risk loci to 101. **b**, Using the 100 RA risk loci (outside of the MHC region), we conducted trans-ethnic and functional annotation of the RA risk SNPs. We constructed an *in silico* bioinformatics pipeline to prioritize biological candidate genes. We adopted eight criteria to score each of 377 genes in the RA risk loci: (1) RA risk missense variant; (2) *cis*-eQTL; (3) PubMed text mining; (4) PPI; (5) PID; (6) haematological cancer somatic mutation; (7) knockout mouse phenotype; and (8) molecular pathway. Our study also demonstrated that these biological candidate genes in RA risk loci are significantly enriched in overlap with target genes for approved RA drugs.



Extended Data Figure 2 | Quantile–quantile plots and Manhattan plots of *P* values in the GWAS meta-analysis. **a**, Quantile–quantile plots of *P* values in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries. The *x*-axis indicates the expected $-\log_{10}$ (*P* values). The *y*-axis indicates the observed $-\log_{10}$ (*P* values) after the application of double GC correction. The SNPs for which observed *P* values were less than 1.0×10^{-20} are indicated at the upper limit of each plot. Black, blue and red dots represent the association results of all SNPs, SNPs outside of the MHC region and *PTPN22* locus, and SNPs outside of the known RA risk loci, respectively. Double GC correction was applied based on the inflation factor, λ_{GC}, which was estimated from the SNPs outside of the known RA loci and indicated in each plot. **b**, Manhattan plots of *P* values in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries. The *y*-axis indicates the $-\log_{10}$ (*P* values) of genome-wide SNPs in each GWAS meta-analysis. The horizontal grey line represents the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$. The SNPs for which *P* values were less than 1.0×10^{-20} are indicated at the upper limit of each plot.







d

Cell types	*P* for H3K4me3 enrichment
T_{reg} primary cells	$\leq 1.0\times10^{-5}$
$CD4^+$ memory primary cells	3.0×10^{-5}
$CD4^+$ naive primary cells	0.0041
$CD8^+$ memory primary cells	0.0065
Smooth muscle, rectal	0.034
Mucosa, colon	0.038
$CD8^+$ naive primary cells	0.12
Mucosa, stomach	0.13
$CD34^+$ primary cells	0.18
$CD34^+$ cultured cells	0.19
Mobilized $CD34^+$ primary cells	0.19
$CD19^+$ primary cells	0.24
$CD3^+$ primary cells	0.30
Mucosa, duodenum	0.40
Muscle satellite cultured cells	0.46
Cingulate gyrus (brain)	0.53
Skeletal muscle	0.77
Mucosa, rectal	0.77
Smooth muscle, colon	0.79
Mesenchymal stem cells (adipose)	0.81
Adipose nuclei	0.84
Smooth muscle, duodenum	0.85
Mid frontal lobe (brain)	0.86
Hippocampus middle (brain)	0.91
Mesenchymal stem cells (bone marrow)	0.91
Pancreatic islets	0.93
Inferior temporal lobe (brain)	0.93
Substantia nigra (brain)	0.93
Adult kidney	0.94
Adult liver	0.95
Mesenchymal stem cells (adipocyte)	0.98
Mesenchymal stem cells (chondrocytes)	0.99
Anterior caudate (brain)	0.99
Smooth muscle, stomach	0.99

e

31 RA risk loci with $P < 10^{-3}$
in EUR and ASN meta-analysis

21.9 SNPs/locus in LD (r^2>0.8) in EUR

37.3 SNPs/locus in LD (r^2>0.8) in ASN

15.0 SNPs/locus in LD (r^2>0.8) in EUR and ASN

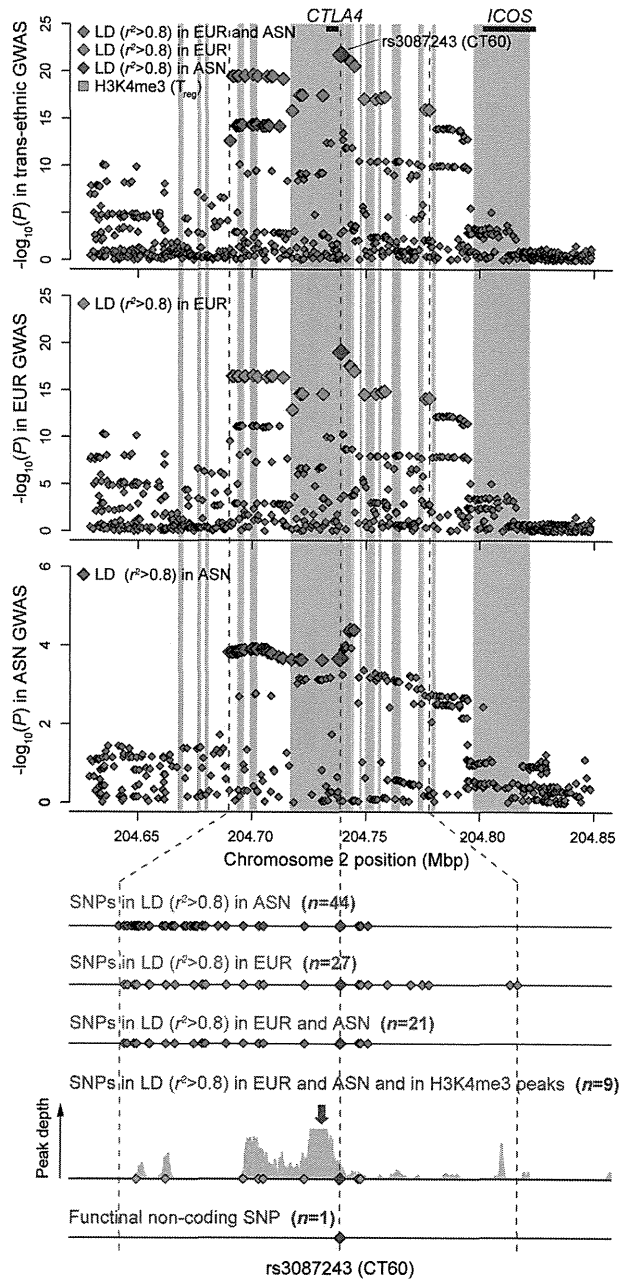
⇩

10 RA risk loci significantly enriched with
T_{reg} primary cell H3K4me3 peaks ($P < 0.05$)

10.4 SNPs/locus in LD (r^2>0.8) in EUR and ASN

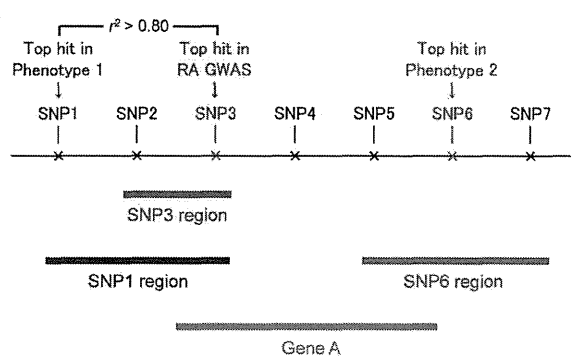
5.9 SNPs/locus in LD (r^2>0.8) in EUR and ASN
and overlapping with H3K4me3 peaks

f



Extended Data Figure 3 | Trans-ethnic and functional annotation of RA risk SNPs. **a**, **b**, Comparisons of RAF and OR values between individuals of European (EUR) and Asian (ASN) ancestry from the stage 1 GWAS meta-analysis. ORs were defined based on minor alleles in Europeans. SNPs with $F_{ST} > 0.10$ or SNPs in which the 95% CI of the OR did not overlap between Europeans and Asians are coloured. OR of the SNP in the *HLA-DRB1* locus (≥ 1.5) is plotted at the upper limits of the *x*- and *y*-axes. Five loci demonstrated population-specific associations ($P < 5.0 \times 10^{-8}$ in one population but $P > 0.05$ in the other population without overlap of the 95% CI of the OR) are highlighted by red labels (rs227163 at *TNFRSF9*, rs624988 at *CD2*, rs726288 at *SFTPD*, rs10790268 at *CXCR5* and rs73194058 at *IFNGR2*). **c**, Cumulative curve of explained heritability in each population. **d**, Enrichment analysis for overlap of RA risk SNPs with H3K4me3 peaks in cell types. The most significant cell type is T_{reg} primary cells. **e**, Number of SNPs in the process of trans-ethnic and functional fine mapping. For 31 loci in which the risk SNPs yielded $P < 1.0 \times 10^{-3}$ in both populations (stage 1 GWAS), the number of candidate causal variants was reduced by 40–70% when confined by SNPs in linkage disequilibrium with the RA risk SNPs ($r^2 > 0.80$) in both populations (on average, from 21.9 or 37.3 SNPs in linkage disequiliberium in Europeans or Asians, to 15.0 SNPs in linkage disequilibrium in both populations). Further, for 10 loci in which candidate causal variants significantly overlapped with H3K4me3 peaks in T_{reg} cells ($P < 0.05$), the average number of SNPs was further reduced by half again, from 10.4 to 5.9. **f**, Fine mapping in the *CTLA4* locus, where the functional non-coding variant of CT60 (rs3087243)[28] showed the most significant association with RA. The top three panels indicate regional SNP associations of the locus in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries, respectively. The bottom panel indicates the change in the number of the candidate causal variants in each process of fine mapping. Trans-ethnic fine mapping of candidate causal variants decreased the number of candidate variants from 44 (linkage disequilibrium in Asians) and 27 (linkage disequilibrium in Europeans) to 21 (linkage disequilibrium in both populations). As these SNPs were significantly enriched in overlap with H3K4me3 peaks in T_{reg} cells compared with the surrounding SNPs ($P = 0.037$), we confined the candidate variants into nine by additionally selecting the SNPs included in H3K4me3 peaks. CT60 was included in these finally selected nine SNPs, and also located at the vicinity of a H3K4me3 peak summit (indicated by a red arrow).

a



RA and Phenotype 1 : Both region-based and allele-based pleiotropy.

RA and Phenotype 2 : Region-based pleiotropy only.

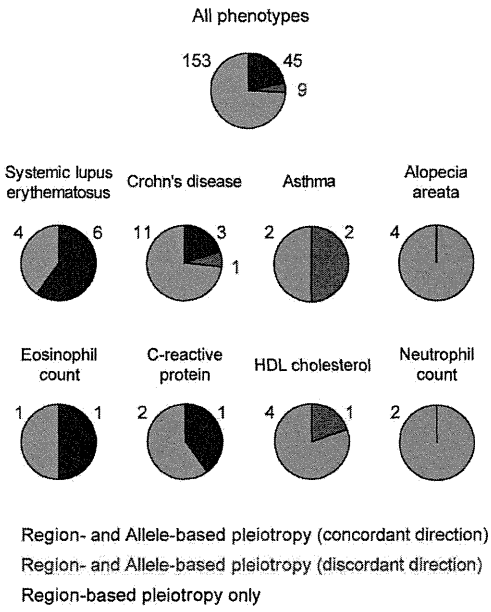
b

Phenotype in GWAS catalogue	No. loci	Region-based pleiotropy		Allele-based pleiotropy
		No. overlap	P-value	
Type 1 diabetes	42	15	$<1.0\times10^{-7}$	7
Crohn's disease	79	15	$<1.0\times10^{-7}$	4
Systemic lupus erythematosus	22	10	$<1.0\times10^{-7}$	6
Celiac disease	26	10	$<1.0\times10^{-7}$	3
Vitiligo	23	9	$<1.0\times10^{-7}$	3
Primary biliary cirrhosis	22	7	2.4×10^{-6}	3
Alopecia areata	5	4	4.5×10^{-6}	0
Ulcerative colitis	52	9	2.5×10^{-5}	3
Multiple sclerosis	52	9	2.5×10^{-5}	2
Chronic lymphocytic leukemia	9	4	9.1×10^{-5}	0
Kawasaki disease	5	3	2.4×10^{-4}	2
Graves' disease	5	3	2.4×10^{-4}	1
Systemic sclerosis	5	3	2.4×10^{-4}	1
Fibrinogen	8	3	0.0012	1
Asthma	17	4	0.0015	2
Psoriasis	18	4	0.0019	1
Hypothyroidism	4	2	0.0041	2
Basal cell carcinoma	5	2	0.0069	0
Neutrophil count	5	2	0.0069	0
HDL cholesterol	46	5	0.014	1
Eosinophil counts	8	2	0.018	1
C-reactive protein	20	3	0.020	1
Melanoma	11	2	0.034	0
Myasthenia gravis	2	1	0.039	1
Primary sclerosing cholangitis	2	1	0.039	0
Soluble ICAM-1	2	1	0.039	0

c

SNP	Chr.	Position (bp)	A1/A2	Gene	Phenotype	Direction
chr1:2523811	1	2,523,811	G/A	*TNFRSF14-MMEL1*	Multiple sclerosis	Concordant
rs2476601	1	114,377,568	A/G	*PTPN22*	Hypothyroidism	Concordant
					Myasthenia gravis	Concordant
					Crohn's disease	Discordant
					Type 1 diabetes	Concordant
rs2228145	1	154,426,970	A/C	*IL6R*	C-reactive protein	Concordant
					Asthma	Discordant
					sIL-6R	Discordant
					Fibrinogen	Concordant
rs2317230	1	157,674,997	T/G	*FCRL3*	Graves' disease	Concordant
rs34695944	2	61,124,850	C/T	*REL*	Hodgkin lymphoma	Concordant
					Psoriasis	Discordant
rs11889341	2	191,943,742	T/C	*STAT4*	Systemic sclerosis	Concordant
					Systemic lupus erythematosus	Concordant
rs3087243	2	204,738,919	G/A	*CTLA4*	Type 1 diabetes	Concordant
rs11933540	4	26,120,001	C/T	*C4orf52*	Type 1 diabetes	Concordant
rs17264332	6	138,005,515	G/A	*TNFAIP3*	Celiac disease	Concordant
					Ulcerative colitis	Concordant
rs7752903	6	138,227,364	G/T	*TNFAIP3*	Systemic lupus erythematosus	Concordant
chr7:128580042	7	128,580,042	G/A	*IRF5*	Ulcerative colitis	Concordant
					Systemic lupus erythematosus	Concordant
rs2736337	8	11,341,880	C/T	*BLK*	Kawasaki disease	Concordant
					Systemic lupus erythematosus	Concordant
rs1516971	8	129,542,100	T/C	*PVT1*	Ovarian cancer	Concordant
					Crohn's disease	Concordant
rs947474	10	6,390,450	A/G	*PRKCQ*	Type 1 diabetes	Concordant
rs2671692	10	50,097,819	A/G	*WDFY4*	Systemic lupus erythematosus	Concordant
rs726288	10	81,706,973	T/C	*SFTPD*	Serum SP-D levels	Concordant
rs4409785	11	95,311,422	C/T	*CEP57*	Vitiligo	Concordant
rs10790268	11	118,729,391	G/A	*CXCR5*	Primary biliary cirrhosis	Concordant
rs61432431	11	128,322,622	C/T	*ETS1*	Systemic lupus erythematosus	Concordant
rs773125	12	56,394,954	A/G	*CDK2*	Polycystic ovary syndrome	Discordant
					Vitiligo	Discordant
					Type 1 diabetes	Discordant
rs10774624	12	111,833,788	G/A	*SH2B3-PTPN11*	Eosinophil counts	Concordant
					Hypothyroidism	Concordant
					Platelet-related traits	Concordant
					Type 1 diabetes	Concordant
					Blood pressure and hypertension	Concordant
					Vitiligo	Concordant
					Retinal vascular caliber	Concordant
					CKD	Concordant
					Celiac disease	Concordant
rs1950897	14	68,760,141	T/C	*RAD51B*	Primary biliary cirrhosis	Concordant
rs13330176	16	86,019,087	A/T	*IRF8*	Multiple sclerosis	Concordant
chr17:38031857	17	38,031,857	G/T	*IKZF3-CSF3*	Primary biliary cirrhosis	Concordant
					Ulcerative colitis	Concordant
					Crohn's disease	Concordant
					Asthma	Discordant
					Type 1 diabetes	Concordant
rs4239702	20	44,749,251	C/T	*CD40*	Kawasaki disease	Concordant
rs2236668	21	45,650,009	C/T	*ICOSLG-AIRE*	Celiac disease	Concordant
rs11089637	22	21,979,096	C/T	*UBE2L3-YDJC*	Crohn's disease	Concordant
					HDL	Discordant

d

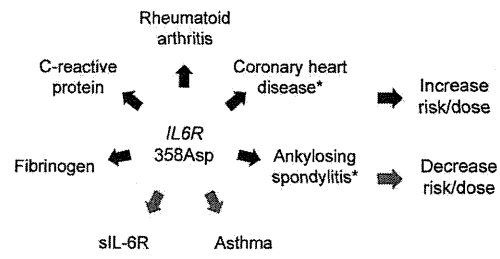


Region- and Allele-based pleiotropy (concordant direction)

Region- and Allele-based pleiotropy (discordant direction)

Region-based pleiotropy only

e

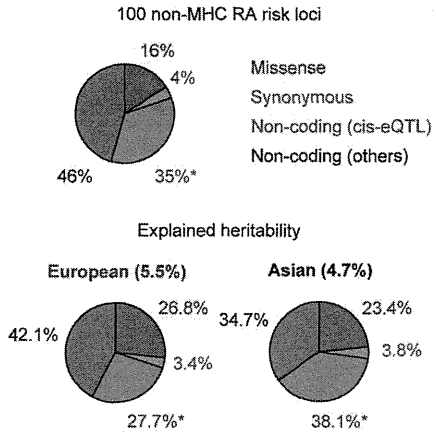


Extended Data Figure 4 | Pleiotropy of RA risk SNPs. **a**, Definition of region-based and allele-based pleiotropy. For each of the RA risk SNPs and SNPs registered in the NHGRI GWAS catalogue (outside of the MHC region), we defined the region on the basis of ±25 kb of the SNP or the neighbouring SNP positions in moderate linkage disequilibrium with it in Europeans or Asians ($r^2 > 0.50$). We defined 'region-based pleiotropy' as two phenotype-associated SNPs sharing part of their genetic regions or any UCSC hg19 reference gene(s) partly overlapping with each of the regions. We defined 'allele-based pleiotropy' as two phenotype-associated SNPs in linkage disequilibrium in Europeans or Asians ($r^2 > 0.80$). **b**, Region-based pleiotropy of the RA risk loci. We found two-thirds of RA risk loci ($n = 66$) demonstrated region-based pleiotropy with other human phenotypes. Phenotypes which showed region-based pleiotropy with RA risk loci are indicated ($P < 0.05$). **c**, Allele-based pleiotropy of the RA risk loci. Allele-based pleiotropy with discordant directional effects to RA risk SNPs are indicated in grey. **d**, Relative proportions of pleiotropic effects (that is, regions and alleles that influence multiple phenotypes) between RA risk loci and 311 phenotypes from the NHGRI GWAS catalogue. Representative examples of disease and biomarker phenotypes are shown. One-quarter of the observed region-based pleiotropic associations (26% = 54/207) were also annotated as having allele-based pleiotropy, although their proportions and directional effects varied among phenotypes. **e**, Allele-based pleiotropy of *IL6R* 358Asp (rs2228145 (A))[5] on multiple disease phenotypes, including increased risk of RA, ankylosing spondylitis and coronary heart disease (asterisks indicate associations obtained from the literature[29,30]) and protection from asthma, as well as levels of biomarkers (increased C-reactive protein (CRP) and fibrinogen but decreased soluble interleukin-6 receptor (sIL6R)).

a

RA risk SNP	r^2	Gene	Missense variants
rs2301888	0.95	*PADI4*	Gly55Ser, Val82Ala, Gly112Ala
rs2476601	1.00	*PTPN22*	Arg620Trp
rs2228145	1.00	*IL6R*	Asp358Ala
rs9826828	0.92	*NCK1*	Ala116Val
	1.00	*NFKBIE*	Val194Ala, Pro175Leu
rs2233424	0.94	*TCTE1*	Arg59His
	0.88	*AARS2*	Val730Met
rs7752903	1.00	*TNFAIP3*	Phe127Cys
rs2671692	0.84	*WDFY4*	Arg1816Gln
rs6479800	0.88	*RTKN2*	Ala288Thr
rs508970	0.90	*CD5*	Ala471Val
rs10774624	0.86	*SH2B3*	Trp262Arg
rs3783782	1.00	*PRKCH*	Val374Ile
rs2582532	1.00	*AHNAK2*	Gly1901Ser
chr17:38031857	0.99	*ZPBP2*	Ser151Ile
	0.99	*GSDMB*	Pro298Ser, Gly291Arg
rs34536443	0.87	*TYK2*	Pro1104Ala
rs2236668	0.94	*ICOSLG*	Trp353Arg
rs5987194	0.96	*IRAK1*	Phe196Ser, Ser453Leu

b



c

	PID classification	No. PID genes	No. overlap with RA genes	Overlap genes	P-value
	All PID genes	194	14	-	1.2×10^{-4}
I	Combined immunodeficiencies	43	3	*PTPRC, RAG1/2, CD40*	0.046
II	Well-defined syndromes	25	2	*ATM, TYK2*	0.12
III	Primary antibody deficiencies	21	2	*CD40, UNG*	0.030
IV	Immune dysregulation	21	4	*CASP8, CASP10, AIRE, IL2RA*	0.0033
V	Phagocyte defects	33	2	*IFNGR2, IRF8*	0.16
VI	Innate immunity	19	0	-	1.0
VII	Autoinflammatory	13	1	*MVK*	0.16
VIII	Complement deficiencies	27	1	*C5*	0.33

d

Cancer type	No. cancer somatic mutation genes	No. overlap with RA genes	Overlap genes	P-value
All cancers	444	23	-	4.7×10^{-5}
Hematological cancers	251	17	-	1.2×10^{-4}
Non-hematological cancers	221	6	-	0.56
Hodgkin lymphoma	10	2	*REL, TNFAIP3*	0.010
B cell non-Hodgkin lymphoma	8	2	*DDX6, FCRL4*	0.015
Non-Hodgkin lymphoma	21	2	*FGFR1OP, HSP90AB1*	0.067
Acute lymphocytic leukemia	29	3	*FCGR2B, AFF3, CDK6*	0.079
Acute myelogenous leukemia	68	2	*ACSL6, PTPN11*	0.47

e

Konckout mouse phenotype category	No. kockout mouse genes with human ortholog	No. overlap with RA genes	P-value
Hematopoietic system phenotype	**2,159**	**86**	**7.0×10^{-6}**
Immune system phenotype	**2,622**	**94**	**1.2×10^{-5}**
Cellular phenotype	**2,961**	**97**	**0.0015**
Liver/biliary system phenotype	982	35	0.0091
Renal/urinary system phenotype	1,028	35	0.011
Endocrine/exocrine gland phenotype	1,453	45	0.020
Respiratory system phenotype	1,097	31	0.028
Tumorigenesis	807	30	0.049
Normal phenotype	1,599	42	0.18
Homeostasis/metabolism phenotype	3,356	88	0.20
Integument phenotype	1,455	35	0.27
Pigmentation phenotype	355	9	0.31
Cardiovascular system phenotype	1,987	42	0.51
Skeleton phenotype	1,435	34	0.57
Other phenotype	258	6	0.57
No phenotypic analysis	1,053	21	0.59
Mortality/aging	3,952	93	0.75
Adipose tissue phenotype	617	12	0.78
Growth/size phenotype	3,061	67	0.79
Digestive/alimentary phenotype	1,128	22	0.80
Reproductive system phenotype	1,730	37	0.81
Limbs/digits/tail phenotype	748	13	0.82
Taste/olfaction phenotype	123	1	0.85
Hearing/vestibular/ear phenotype	557	8	0.88
Embryogenesis phenotype	1,535	30	0.92
Behavior/neurological phenotype	2,465	46	0.94
Nervous system phenotype	2,805	53	0.95
Craniofacial phenotype	951	15	0.96
Muscle phenotype	1,198	21	0.96
Vision/eye phenotype	1,214	21	0.99

f

Database	Molecular pathway	Pathway enrichment (FDR q) Current study	Previous study
BIOCARTA	B Lymphocyte Cell Surface Molecules	**2.0×10^{-4}**	0.26
BIOCARTA	T Cytotoxic Cell Surface Molecules	**3.3×10^{-4}**	**0.032**
BIOCARTA	T Helper Cell Surface Molecules	**4.0×10^{-4}**	**0.030**
BIOCARTA	Th1/Th2 Differentiation	**0.0025**	**0.0063**
Ingenuity	IL-10.Signaling	**0.0026**	0.46
Ingenuity	Interferon.Signaling	**0.0028**	0.13
Ingenuity	GM-CSF.Signaling	**0.0031**	0.43
Ingenuity	T.Cell.Receptor.Signaling	**0.0034**	**0.029**
BIOCARTA	NO2-dependent IL 12 Pathway in NK cells	**0.0044**	0.06
BIOCARTA	IL-22 Soluble Receptor Signaling	**0.0046**	0.39
BIOCARTA	The Co-Stimulatory Signal During T-cell Activation	**0.0046**	0.06
BIOCARTA	Selective expression of chemokine receptors during T-cell polarization	**0.0048**	0.21
Ingenuity	Hepatic.Fibrosis.Hepatic.Stellate.Cell.Activation	**0.0073**	**0.0060**
Ingenuity	p38.MAPK.Signaling	**0.0076**	0.19
Ingenuity	Neuregulin.Signaling	**0.0079**	0.51
Ingenuity	IL-6.Signaling	**0.0082**	0.11
Ingenuity	Glucocorticoid.Receptor.Signaling	**0.0090**	0.18
BIOCARTA	IL-6 signaling	**0.0091**	0.50
BIOCARTA	Influence of Ras and Rho proteins on G1 to S Transition	**0.016**	0.38
BIOCARTA	IL-3 signaling	**0.018**	0.64
BIOCARTA	Adhesion and Diapedesis of Granulocytes	**0.018**	0.15
BIOCARTA	RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	**0.018**	0.15
Ingenuity	Fc.Epsilon.RI.Signaling	**0.022**	0.19
Ingenuity	JAK.Stat.Signaling	**0.023**	0.48
Ingenuity	IL-2.Signaling	**0.026**	0.17
Ingenuity	PPAR.Signaling	**0.026**	0.24
BIOCARTA	IL-2 Receptor Beta Chain in T cell Activation	**0.027**	0.39
BIOCARTA	Cyclins and Cell Cycle Regulation	**0.028**	0.16
Ingenuity	Leukocyte.Extravasation.Signaling	**0.028**	0.45
BIOCARTA	p53 Signaling Pathway	**0.028**	0.40
BIOCARTA	Role of ERBB2 in Signal Transduction and Oncology	**0.028**	0.51
Ingenuity	B.Cell.Receptor.Signaling	**0.028**	0.45
BIOCARTA	CD40L Signaling	**0.029**	0.16
BIOCARTA	Cells and Molecules involved in local acute inflammatory response	**0.034**	0.40
BIOCARTA	Antigen Dependent B Cell Activation	**0.036**	0.06
BIOCARTA	Adhesion and Diapedesis of Lymphocytes	**0.043**	0.60
BIOCARTA	MAPKinase Signaling	**0.044**	0.76
BIOCARTA	Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase	**0.044**	0.59
Ingenuity	NFKB.Signaling	**0.045**	0.05
Ingenuity	Aryl.Hydrocarbon.Receptor.Signaling	**0.048**	0.33
Ingenuity	PDGF.Signaling	**0.049**	0.30

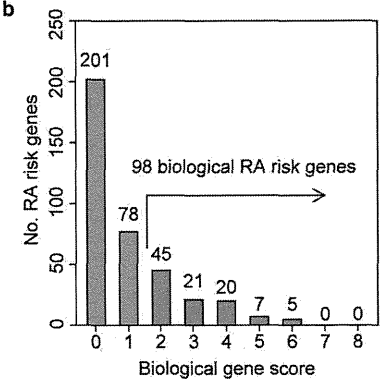
Extended Data Figure 5 | Overlap of RA risk SNPs with biological resources. **a**, Missense variants in linkage disequilibrium ($r^2 > 0.80$ in Europeans or Asians) with RA risk SNPs. When multiple missense variants are in linkage disequilibrium with the RA risk SNP, the highest r^2 value is indicated. **b**, Functional annotation of the SNPs in 100 non-MHC RA risk loci, including the relative proportion of heritability explained by SNP annotations. Although 44% of all RA risk SNPs had *cis*-eQTL, 9 of them overlapped with missense or synonymous variants but 35 of them did not overlap as indicated by asterisks. A list of *cis*-eQTL SNPs and genes can be found in Extended Data Table 2. **c**, Overlap of RA risk genes with human PID and defined categories. **d**, Overlap of RA risk genes with cancer somatic mutation genes. In addition to the categories of all cancers, haematological cancers and non-haematological cancers, cancer types that showed overlap with ≥2 of RA risk genes are indicated. **e**, Overlap of RA risk genes with knockout mouse phenotypes. Knockout mouse phenotypes that satisfied significant enrichment with RA risk genes are indicated in bold ($P < 0.05/30 = 0.0017$). **f**, Molecular pathway analysis of RA GWAS results. Molecular pathways that showed significant enrichment in either the current stage 1 trans-ethnic GWAS meta-analysis or the previous GWAS meta-analysis of RA[2] are indicated in bold (FDR $q < 0.05$).

a

Biological RA risk gene prioritization criteria

(1) RA risk missense variant (n = 19)
(2) Cis-eQTL (n = 51)
(3) PubMed text-mining (n = 90)
(4) Protein-protein interaction (n = 63)
(5) Primary immunodeficiency (n = 15)
(6) Hematological cancer (n = 17)
(7) Knockout mouse phenotype (n = 86)
(8) Molecular pathway (n = 35)

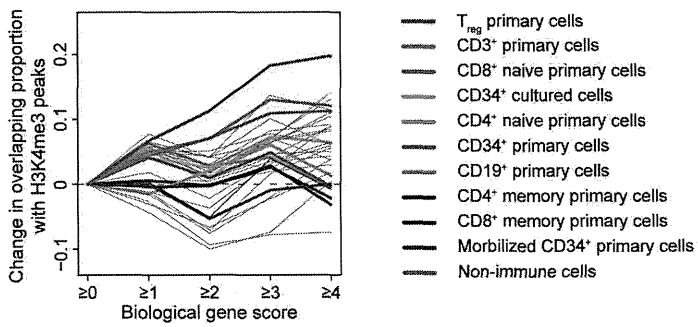
b



c

Correlation of prioritization criteria of biological genes from RA risk loci (R^2)	RA risk missense variant	Cis-eQTL	PubMed text-mining	Protein-protein interaction	Primary immunodeficiency	Hematological cancer	Knockout mouse phenotype	Molecular pathway
RA risk missense variant	-	0.01	0.03	0.02	0.00	0.00	0.02	0.01
Cis-eQTL	0.01	-	0.05	0.01	0.01	0.00	0.02	0.01
PubMed text-mining	0.03	0.05	-	0.10	0.06	0.03	0.26	0.14
Protein-protein interaction	0.02	0.01	0.10	-	0.04	0.01	0.07	0.06
Primary immunodeficiency	0.00	0.01	0.06	0.04	-	0.00	0.08	0.07
Hematological cancer	0.00	0.00	0.03	0.01	0.00	-	0.03	0.04
Knockout mouse phenotype	0.02	0.02	0.26	0.07	0.08	0.03	-	0.21
Molecular pathway	0.01	0.01	0.14	0.06	0.07	0.04	0.21	-

d

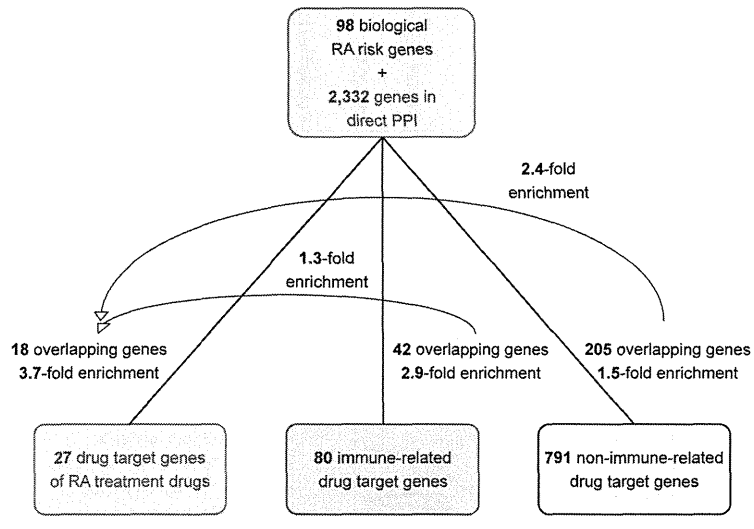


Extended Data Figure 6 | Prioritization of biological candidate genes from RA risk loci. **a**, Prioritization criteria of biological candidate genes from RA risk loci. **b**, Histogram distribution of gene scores. The 98 genes with score ≥2 (orange) were defined as 'biological RA risk genes'. **c**, Correlations of biological candidate gene prioritization criteria. **d**, Change in the overlapping proportions of genes with H3K4me3 peaks by cell type according to score increases. When RA risk SNP of the locus (or SNP in linkage disequilibrium) overlapped with H3K4me3 peaks, genes in the locus were defined as overlapping.

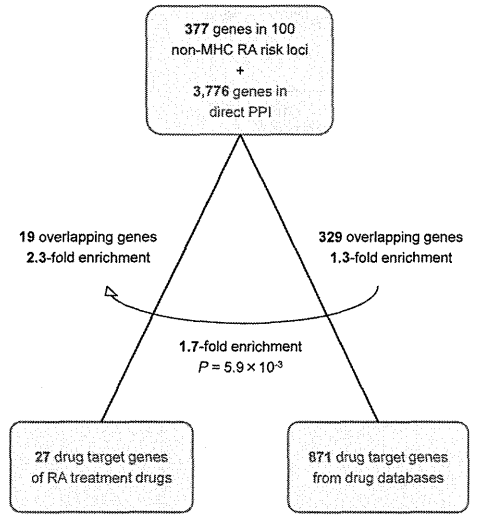
a

RA drug category	Generic name	Target gene
Biologics	Etanercept	
	Infliximab	
	Adalimumab	*TNF*
	Golimumab	
	Certolizumab pegol	
	Abatacept	*CD80, CD86*
	Anakinra	*IL1R1*
	Rituximab	*MS4A1*
	Tocilizumab	*IL6R*
DMARDs	Auranofin	*PRDX5, IKBKB*
	Azathioprine	*HPRT1*
	Cyclophosphamide	-
	Cyclosporine	*CAMLG, PPP3R2*
	Iguratimod (T-614)	*ELANE, PTGS2*
	Leflunomide	*DHODH*
	Methotrexate	*DHFR*
	Sulfasalazine	*ALOX5, PTGS1, PTGS2, PPARG*
	Tacrolimus	*FKBP1A*
	Temsirolimus	*MTOR*
	Tofacitinib (CP-690,550)	*JAK1, JAK2, JAK3*
Steroids	Prednisolone	*NR3C1*
	Methylprednisolone	*NR3C1*
	Desoxycorticosterone Pivalate	*NR3C2*
Others	Hydroxychloroquine	*TLR7, TLR9*

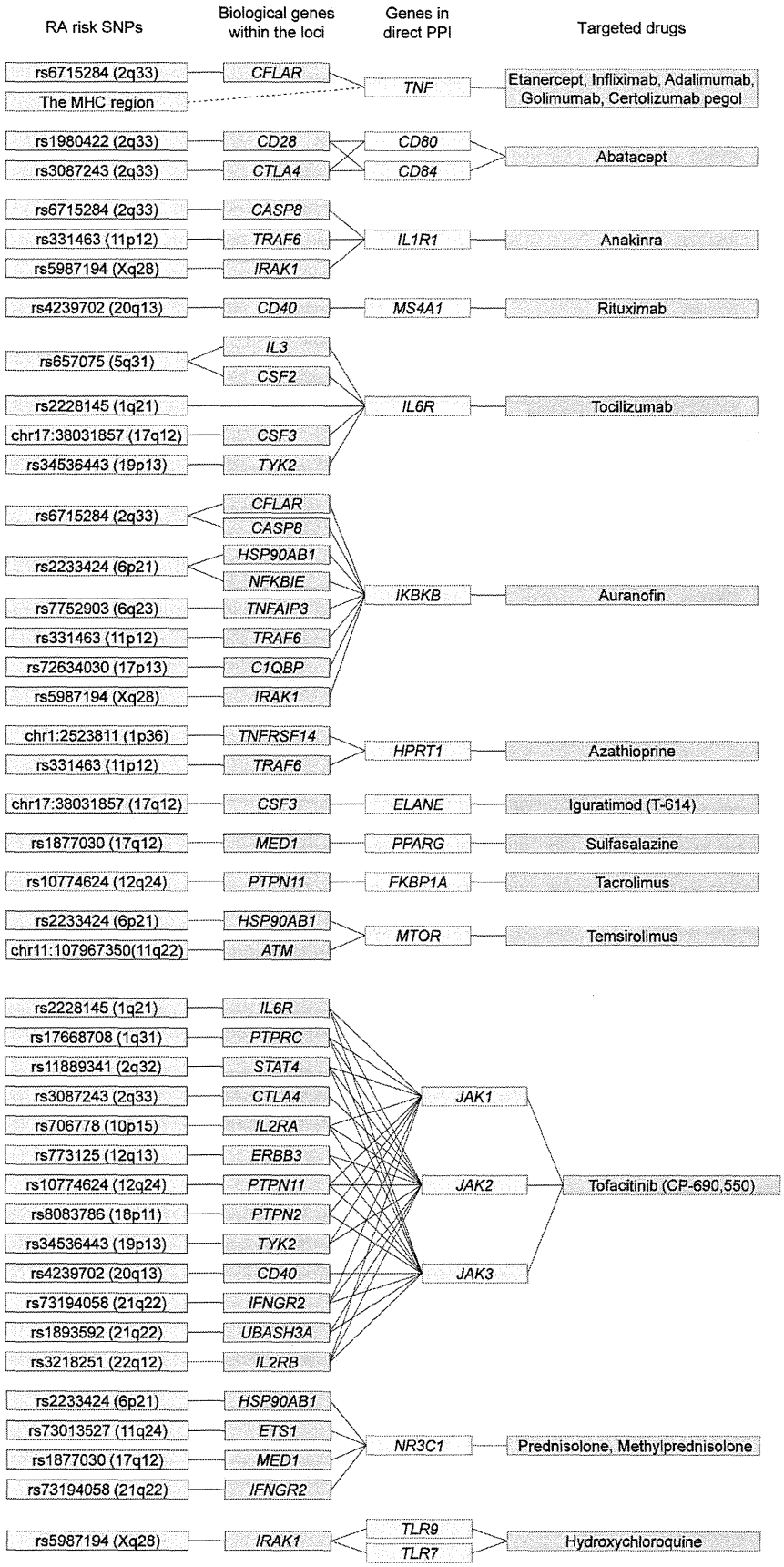
b



c



Extended Data Figure 7 | Overlap of all genes in the RA risk loci with drug target genes. **a**, Approved RA drugs and target genes. DMARDs, disease-modifying antirheumatic drugs. **b**, Overlap analysis stratified by immune-related and non-immune-related drug target genes. We made a list of 583 immune-related genes based on Gene Ontology (GO) pathways named 'immune-' or 'immuno-' and found that the majority of drug target genes (791/871 = 91%) were not immune-related. **c**, Overlap of all 377 genes included in 100 RA risk loci (outside of the MHC region) plus 3,776 genes in direct PPI with them and drug target genes. We found overlap of 19 genes from the 27 drug target genes of approved RA drugs (2.3-fold enrichment, $P < 1.0 \times 10^{-5}$). All 871 drug target genes (regardless of disease indication) overlap with 329 genes from the PPI network, which is 1.3-fold more enrichment than expected by chance alone ($P < 1.0 \times 10^{-5}$), but less than 1.7-fold enrichment compared with RA drugs ($P = 0.0059$). We note that this enrichment of drug–gene pairs was less apparent compared with that obtained from the expanded PPI network generated from 98 biological candidate genes (Fig. 3b).



Extended Data Figure 8 | Connection between RA risk genes and approved RA drugs. Full lists of the connections between RA risk SNPs (blue boxes), biological candidate genes from each risk locus (purple boxes), genes from the expanded PPI network (green boxes) and approved RA drugs (orange boxes). Black lines indicate connections. Only *IL6R* is a direct connection between an SNP–biological gene–drug (tocilizumab)[19,20]; all other SNP–drug connections are through the PPI network.

Extended Data Table 1 | Characteristics of the study cohorts

a

Study stage	Cohort	Ethnicity	Geographical origin	No. subjects			RA case sero-positivity
				Cases	Controls	Total	
GWAS meta-analysis (Stage 1)	BRASS	European	North America	483	1,631	2,114	100% CCP+
	CANADA		Canada	589	1,554	2,143	100% CCP+
	EIRA		Sweden	1,097	1,044	2,141	100% CCP+
	NARAC1		North America	863	1,191	2,054	100% CCP+
	NARAC2		North America	896	6,603	7,499	100% CCP+
	WTCCC		United Kingdom	1,520	10,507	12,027	100% CCP+ or RF+
	RACI-UK		United Kingdom	1,645	6,082	7,727	100% CCP+
	RACI-US		North America	997	2,132	3,129	100% CCP+
	RACI-SE-E		Sweden	740	1,117	1,857	100% CCP+
	RACI-SE-U		Sweden	522	962	1,484	100% CCP+
	RACI-NL		Netherland	303	2,001	2,304	100% CCP+
	RACI-ES		Spain	397	399	796	100% CCP+
	RACI-i2b2		North America	882	1,863	2,745	100% CCP+
	ReAct		France	275	804	1,079	70% CCP+ or RF+
	Dutch (AMC, BeSt, LUMC, DREAM)		Netherland	1,172	1,684	2,856	80% CCP+ or RF+
	ACR-REF (BRAGGSS, BRAGGSS2, ERA, KI, TEAR)		North America & Europe	347	264	611	85% CCP+ or RF+
	CORRONA		North America	894	1,838	2,732	61% CCP+ or RF+, 32% unknown
	Vanderbilt		North America	739	2,247	2,986	31% CCP+ or RF+, 56% unknown
	GARNET (BioBank Japan Project; BBJ)	Asian	Japan	2,414	14,245	16,659	79% CCP+, 76% RF+
	GARNET (Kyoto University)		Japan	1,237	2,087	3,324	85% CCP+, 86% RF+
	GARNET (IORRA)		Japan	423	559	982	87% CCP+, 88% RF+
	Korea		Korea	799	751	1,550	100% CCP+
	European	-	-	14,361	43,923	58,284	-
	Asian	-	-	4,873	17,642	22,515	-
	Trans-ethnic	-	-	19,234	61,565	80,799	-
In-silico replication study (Stage 2)	Genentech	European	North America	2,780	4,700	7,480	44% CCP+, 52% unknown 81% RF+, 1.7% unknown
	China	Asian	China	928	835	1,763	48% CCP+
	Total	-	-	3,708	5,535	9,243	-
De-novo replication study (Stage 3)	CANADAII	European	Canada	995	1,101	2,096	100% CCP+
	GARNET	Asian	Japan	5,943	5,557	11,500	81% CCP+, 86% RF+
	Total	-	-	6,938	6,658	13,596	-
Total	European	-	-	18,136	49,724	67,860	-
	Asian	-	-	11,744	24,034	35,778	-
	Trans-ethnic	-	-	29,880	73,758	103,638	-

b

Study stage	Cohort	Genotyping platform	GWAS QC criteria				Imputation method			No.SNPs after QC		Inflation factor		Covariates	X chrom. data
			Sample call rate	SNP call rate	MAF	HWE P-value	Reference panel	MAF	Quality score	Genotyped	Imputed	λ_{GC}	λ_{GC_1000}		
GWAS meta-analysis (Stage 1)	BRASS	Affymetrix Genome-wide Human SNP Array 6.0	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	649,178	8,201,244	1.015	1.008	Top 5 PCs	Available
	CANADA	Illumina HumanCNV370-Duo BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	295,430	7,933,623	1.002	1.001	Top 5 PCs	Available
	EIRA	HumanHap300 BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	298,193	8,163,538	0.991	0.994	Top 5 PCs	N.A.
	NARAC1	Illumina HumanHap550 BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	479,671	8,254,787	1.017	1.012	Top 5 PCs	N.A.
	NARAC2	HumanHap300 BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	261,974	7,733,592	1.023	1.003	Top 5 PCs	N.A.
	WTCCC	Affymetrix Genome-wide Human SNP Array 5.0	>0.99	>0.99	>0.01	$>10^{-5}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	339,790	7,385,370	1.043	1.004	Top 5 PCs	N.A.
	RACI-UK	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	126,740	873,840	1.058	1.008	Top 10 PCs	Available
	RACI-US	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	120,589	843,395	1.031	1.012	Top 10 PCs	Available
	RACI-SE-E	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	124,801	870,585	1.003	1.002	Top 10 PCs	Available
	RACI-SE-U	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	123,998	870,797	0.986	0.988	Top 10 PCs	Available
	RACI-NL	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	124,480	862,815	1.109	1.051	Top 10 PCs	Available
	RACI-ES	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	124,459	859,540	1.081	1.152	Top 10 PCs	Available
	RACI-i2b2	Illumina Immunochip custom array	>0.99	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.7	118,731	829,507	1.003	1.001	Top 10 PCs	Available
	ReAct	Illumina OmniExpress BeadChip Illumina Human 660W-Quad BeadChip	>0.98	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	257,299	7,588,538	0.992	0.991	Top 5 PCs	Available
	Dutch	Illumina Human 660W-Quad BeadChip Illumina HumanHap550 BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	284,884	7,956,686	1.023	1.011	Top 5 PCs	Available
	ACR-REF	Illumina HumanCNV370-Duo BeadChip Illumina OmniExpress BeadChip Illumina Human 660W-Quad BeadChip	>0.95	>0.95	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	234,075	7,593,678	1.026	1.070	Top 5 PCs	Available
	CORRONA	Illumina OmniExpress BeadChip	>0.98	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	552,896	8,400,238	1.001	1.000	Top 5 PCs	Available
	Vanderbilt	Illumina OmniExpress BeadChip	>0.98	>0.99	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	541,143	8,372,666	0.987	0.995	Top 5 PCs	Available
	BBJ	Illumina HumanHap610-Quad BeadChip Illumina HumaHap610-Quad BeadChip	>0.98	>0.99	>0.01	$>10^{-7}$	1000 Genomes Phase I (α) Asians	>0.005	>0.5	477,784	6,874,738	1.038	1.002	-	Available
	Kyoto	Illumina HumanHap550 BeadChip Illumina HumanCNV370-Duo BeadChip	>0.90	>0.95	>0.05	$>10^{-7}$	1000 Genomes Phase I (α) Asians	>0.005	>0.5	227,348	6,254,431	1.099	1.038	-	N.A.
	IORRA	Affymetrix Genome-wide Human SNP Array 6.0	>0.95	>0.98	>0.05	$>10^{-6}$	1000 Genomes Phase I (α) Asians	>0.005	>0.5	465,832	6,567,923	0.992	0.989	-	Available
	Korea	Illumina Human 660W-Quad BeadChip Illumina HumanHap550 BeadChip	>0.90	>0.90	>0.01	$>10^{-6}$	1000 Genomes Phase I (α) Asians	>0.005	>0.5	418,837	6,424,378	1.007	1.007	-	Available
	European	-	-	-	-	-	-	-	-	-	8,747,962	1.073	1.003	-	-
	Asian	-	-	-	-	-	-	-	-	-	6,619,871	1.041	1.005	-	-
	Trans-ethnic	-	-	-	-	-	-	-	-	-	9,739,303	1.072	1.002	-	-
In-silico replication study (Stage 2)	Genentech	Illumina HumanOmni1-Quad_v1-0_B Illumina Humanhap550K	>0.95	>0.95	>0.10	$>10^{-4}$	1000 Genomes Phase I (α) Europeans	>0.005	>0.5	-	-	-	-	Top 5 PCs	N.A.
	China	Affymetrix Genome-wide Human SNP Array 6.0	>0.95	>0.95	>0.05	$>10^{-3}$	1000 Genomes Phase I (α) Asians	>0.005	>0.5	-	-	-	-	Top 5 PCs	N.A.
De-novo replication study (Stage 3)	CANADAII	iPlex genotying system	-	-	-	-	-	-	-	-	-	-	-	-	Available
	GARNET	Taqman genotyping system	-	-	-	-	-	-	-	-	-	-	-	-	Available

a, Characteristics of the cohorts and subjects enrolled in the study. **b**, Genotype and imputation methods of the studies. CCP, anti-citrullinated peptide antibody; chrom, chromosome; N.A., not available; PC, principal component; QC, quality control; RF, rheumatoid factor.

Extended Data Table 2 | *cis*-eQTL of RA risk SNPs

a

RA risk SNP	Chr.	Position (bp)	eQTL gene	Cis-eQTL effect of best proxy SNP				Cis-eQTL effect of top eQTL SNP			
				Proxy SNP	Position (bp)	eQTL *P*	r^2	eQTL SNP	Position (bp)	eQTL *P*	r^2
chr1:2523811	1	2,523,811	*PLCH2*	rs10910099	2,533,552	2.2E-18	0.87	rs2494435	2,359,280	2.6E-45	<0.25
			TNFRSF14	rs2843401	2,528,133	1.1E-28	0.87	rs734999	2,513,216	2.1E-90	0.43
rs227163	1	7,961,206	*PARK7*	rs227163	7,961,206	4.6E-10	1.00	rs3766606	8,022,197	1.0E-53	<0.25
rs28411352	1	38,278,579	*MANEAL,YRDC*	rs2306627	38,260,503	3.9E-09	0.84	rs2306426	36,451,618	7.7E-10	<0.25
			INPP5B	rs2306627	38,260,503	7.5E-23	0.84	rs4072980	38,456,106	1.2E-113	<0.25
			SF3A3	rs2306627	38,260,503	3.3E-17	0.84	rs4072980	38,456,106	1.1E-190	<0.25
			FHL3	rs2306627	38,260,503	1.1E-11	0.84	rs4634868	38,465,315	9.8E-198	<0.25
rs2476601	1	114,377,568	*PTPN22*	rs2476601	114,377,568	3.4E-10	1.00	rs7555634	114,367,116	5.3E-43	<0.25
rs2228145	1	154,426,970	*AQP10*	rs6684439	154,395,839	3.3E-06	0.89	rs6668968	154,293,675	3.8E-40	<0.25
			IL6R	rs4129267	154,426,264	3.2E-27	1.00	rs4537545	154,418,879	2.0E-29	0.96
			UBE2Q1	rs4129267	154,426,264	9.7E-08	1.00	rs6660775	154,538,554	3.9E-21	<0.25
rs2317230	1	157,674,997	*FCRL5*	rs3761959	157,669,278	1.7E-09	0.87	rs6427386	157,530,097	9.8E-198	<0.25
			FCRL3	rs7528684	157,670,816	9.8E-198	0.87	rs2210913	157,668,993	9.8E-198	0.87
rs4656942	1	160,831,048	*LY9*	rs4656942	160,831,048	2.7E-96	1.00	rs576334	160,797,514	5.6E-195	<0.25
rs72717009	1	161,405,053	*SDHC*	rs12731669	161,410,458	5.5E-05	0.97	rs16832871	161,335,758	1.4E-142	<0.25
			FCGR2B	rs12731669	161,410,458	4.2E-83	0.97	rs6674499	161,618,151	9.8E-198	<0.25
rs17668708	1	198,640,488	*PTPRC*	rs17669032	198,653,174	5.2E-05	0.97	rs2296618	198,666,232	2.1E-05	0.78
rs1980422	2	204,610,396	*CD28*	rs1980421	204,610,004	7.3E-18	1.00	rs2140148	204,572,140	8.1E-21	0.47
rs10028001	4	79,502,972	*ANXA3*	rs10028001	79,502,972	1.1E-04	1.00	rs4975144	79,474,040	1.4E-09	<0.25
rs2561477	5	102,608,924	*PAM*	rs411648	102,602,902	2.2E-113	1.00	rs2431321	102,118,794	9.8E-198	<0.25
			GIN1	rs2288786	102,600,754	1.3E-06	1.00	rs42431	102,400,063	2.6E-13	0.43
rs657075	5	131,430,118	*ACSL6*	rs657075	131,430,118	3.8E-12	1.00	rs253946	131,330,461	9.2E-26	0.32
chr6:14103212	6	14,103,212	*CD83*	rs12530098	14,107,197	2.6E-24	1.00	rs16874672	14,087,484	2.2E-26	0.90
			KCTD20	rs4713969	36,349,008	8.2E-05	0.99	rs4711453	36,439,391	3.1E-32	<0.25
rs2234067	6	36,355,654	*STK38*	rs4713969	36,349,008	1.4E-06	0.99	rs1812018	36,557,976	6.8E-15	<0.25
			-	rs4713969	36,349,008	2.1E-26	0.99	rs10947614	36,573,822	1.1E-146	<0.25
			SFRS3	rs4713969	36,349,008	2.6E-11	0.99	rs7743396	36,579,252	1.5E-52	<0.25
rs9373594	6	149,834,574	*C6orf72*	rs9377224	149,853,707	4.0E-06	1.00	rs9322189	149,909,933	1.8E-15	0.30
			NUP43	rs9377224	149,853,707	4.1E-64	1.00	rs9688350	150,052,113	9.8E-198	0.27
rs2451258	6	159,506,600	*RSPH3*	rs2485363	159,506,121	5.0E-05	1.00	rs12216499	159,368,524	2.0E-119	<0.25
rs1571878	6	167,540,842	*RNASET2*	rs1571878	167,540,842	9.8E-198	1.00	rs429083	167,383,972	9.8E-198	0.39
chr7:128580042	7	128,580,042	*TNPO3*	rs3807306	128,580,680	1.4E-150	0.81	rs3807306	128,580,680	1.4E-150	0.81
			-	rs3807306	128,580,680	2.4E-32	0.81	rs10229001	128,599,397	4.5E-49	0.48
			IRF5	rs3807306	128,580,680	9.8E-198	0.81	rs7807018	128,640,188	9.8E-198	0.48
rs2736337	8	11,341,880	*C8orf13,C8orf12*	rs2736340	11,343,973	1.6E-174	0.99	rs4840568	11,351,019	3.8E-175	0.92
			BLK	rs1478901	11,347,833	1.8E-120	0.99	rs998683	11,353,000	1.5E-120	0.95
rs10985070	9	123,636,121	*TRAF1*	rs10985070	123,636,121	3.9E-72	1.00	rs2416804	123,676,396	3.8E-73	0.96
			PHF19	rs10985070	123,636,121	2.9E-10	1.00	rs10760129	123,700,183	2.2E-10	0.95
			C5	rs10985070	123,636,121	4.9E-68	1.00	rs2416811	123,789,634	2.0E-146	0.31
rs947474	10	6,390,450	-	rs947474	6,390,450	6.5E-06	1.00	rs12416248	6,391,031	1.1E-43	<0.25
rs2671692	10	50,097,819	*WDFY4*	rs2671692	50,097,819	3.0E-09	1.00	rs7072606	49,933,974	1.1E-50	<0.25
rs968567	11	61,595,564	*C11orf10*	rs968567	61,595,564	3.1E-39	1.00	rs174538	61,560,081	2.5E-67	0.40
			FADS1	rs968567	61,595,564	8.1E-62	1.00	rs968567	61,595,564	8.1E-62	1.00
			FADS2	rs968567	61,595,564	4.8E-34	1.00	rs968567	61,595,564	4.8E-34	1.00
rs10774624	12	111,833,788	*SH2B3*	rs653178	112,007,756	1.7E-19	0.86	rs2239195	111,881,309	1.0E-33	<0.25
			ALDH2	rs653178	112,007,756	8.7E-07	0.86	rs16941669	112,245,637	4.4E-50	<0.25
rs4780401	16	11,839,326	*TXNDC11*	rs11075010	11,826,013	8.3E-09	0.93	rs12919035	11,821,508	4.4E-12	0.49
rs72634030	17	5,272,580	*ZNF594*	rs8080217	5,164,761	8.7E-11	0.88	rs2071456	5,031,946	1.5E-12	0.65
			C17orf87	rs8080217	5,164,761	3.3E-05	0.88	rs2641232	5,087,602	1.4E-53	<0.25
			-	rs8080217	5,164,761	3.6E-70	0.88	rs7426	5,288,983	9.8E-198	<0.25
			NUP88	rs8080217	5,164,761	3.3E-27	0.88	rs1989946	5,313,152	8.9E-96	<0.25
			MIS12	rs8080217	5,164,761	8.5E-10	0.88	rs1805448	5,384,327	2.2E-35	<0.25
rs1877030	17	37,740,161	*FBXL20*	rs12937013	37,665,571	3.4E-15	1.00	rs8076462	37,400,025	3.1E-42	<0.25
			PPP1R1B	rs1877030	37,740,161	1.8E-10	1.00	rs879606	37,781,849	8.0E-18	0.41
			-	rs11657058	37,699,378	3.9E-05	1.00	rs7219814	37,478,801	2.1E-111	<0.25
			IKZF3	rs4795385	37,733,148	8.8E-24	1.00	rs2517955	37,843,681	5.2E-82	0.33
chr17:38031857	17	38,031,857	-	rs907092	37,922,259	6.6E-11	0.90	rs7219814	37,478,801	2.1E-111	<0.25
			IKZF3	rs11557467	38,028,634	3.3E-05	0.84	rs9896940	37,895,975	3.1E-25	<0.25
			GSDMB	rs10852936	38,031,714	9.8E-198	0.98	rs9901146	38,043,343	9.8E-198	0.84
			ORMDL3	rs10852936	38,031,714	9.8E-198	0.98	rs8076131	38,080,912	9.8E-198	0.86
rs2469434	18	67,544,046	*CD226*	rs1610555	67,543,147	2.3E-33	0.99	rs763361	67,531,642	2.4E-50	0.66
rs4239702	20	44,749,251	*CD40*	rs4239702	44,749,251	1.3E-34	1.00	rs745307	44,747,086	1.5E-72	<0.25
rs73194058	21	34,764,288	*IL10RB*	rs11702844	34,759,876	1.3E-11	0.97	rs1058867	34,669,381	3.0E-69	<0.25
			IFNAR1	rs11702844	34,759,876	8.0E-12	0.97	rs2257167	34,715,699	4.2E-73	<0.25
			TMEM50B	rs11702844	34,759,876	3.1E-11	0.97	rs1059293	34,809,693	2.2E-103	<0.25
			-	rs11702844	34,759,876	2.8E-34	0.97	rs2834217	34,822,150	9.8E-198	<0.25
rs1893592	21	43,855,067	*UBASH3A*	rs1893592	43,855,067	6.4E-92	1.00	rs1893592	43,855,067	6.4E-92	1.00
rs2236668	21	45,650,009	*ICOSLG*	rs7278940	45,648,992	3.7E-06	1.00	rs3788111	45,668,171	8.4E-16	<0.25
rs11089637	22	21,979,096	-	rs11089637	21,979,096	9.8E-198	1.00	rs5754217	21,939,675	9.8E-198	0.87
rs909685	22	39,747,671	*SYNGR1*	rs909685	39,747,671	1.0E-140	1.00	rs909685	39,747,671	1.0E-140	1.00
			MAP3K7IP1	rs909685	39,747,671	1.3E-05	1.00	rs5750824	39,830,123	5.9E-07	0.28

b

SNP	Chr.	Position (bp)	eQTL gene	Nominal *P* for cis-eQTL	
				$CD4^+$ T-cell	$CD14^+16^-$ Monocyte
rs28411352	1	38,278,579	*INPP5B*	0.022	**3.6E-16**
			FHL3	0.081	**8.9E-13**
rs2317230	1	157,674,997	*FCRL3*	**3.5E-06**	0.87
rs9653442	2	100,825,367	*AFF3*	**5.2E-08**	0.18
rs7731626	5	55,444,683	*IL6ST*	**2.3E-07**	0.0087
			ANKRD55	**4.1E-14**	0.43
rs2234067	6	36,355,654	*ETV7*	2.9E-04	**1.1E-10**
rs9373594	6	149,834,574	*NUP43*	5.4E-04	**1.5E-05**
rs1571878	6	167,540,842	*RNASET2*	**6.9E-20**	**1.3E-05**
rs67250450	7	28,174,986	*JAZF1*	**3.6E-17**	2.0E-04
chr7:128580042*	7	128,580,042	*TNPO3*	1.0E-04	**3.0E-07**
rs10985070	9	123,636,121	*MEGF9*	**3.3E-06**	0.10
			PSMD5	0.017	**1.8E-05**
			PHF19	0.0016	**5.6E-06**
rs968567	11	61,595,564	*FADS2*	**1.4E-31**	**8.9E-35**
			FADS1	**2.1E-32**	0.094
rs11605042	11	72,411,664	*STARD10*	0.82	**1.0E-07**
rs4409785	11	95,311,422	*SESN3*	**1.5E-11**	0.43
rs773125	12	56,394,954	*SUOX*	0.27	**1.1E-09**
rs1633360	12	58,108,052	*TSPAN31*	0.13	**1.0E-05**
			METTL21B	**1.4E-09**	**4.0E-10**
rs9603616	13	40,368,069	*COG6*	0.0011	**1.2E-05**
rs4780401	16	11,839,326	*TXNDC11*	**1.3E-05**	0.62
rs72634030	17	5,272,580	*MIS12*	0.0039	**1.3E-05**
rs1877030	17	37,740,161	*STARD3*	0.048	**4.5E-05**
chr17:38031857†	17	38,031,857	*GSDMA*	**2.1E-06**	0.63
			GSDMB	**4.3E-11**	0.19
			ORMDL3	**6.8E-09**	0.0098
rs4239702	20	44,749,251	*CD40*	0.31	**1.7E-08**
rs73194058	21	34,764,288	*IFNGR2*	0.096	**1.9E-06**
			TMEM50B	**7.5E-07**	0.013
rs1893592	21	43,855,067	*UBASH3A*	**3.8E-14**	0.92

a, *cis*-eQTL of PBMCs in the RA risk SNPs. Significant *cis*-eQTLs of RA risk SNPs is indicated (FDR $q < 0.05$). SNPs and positions are based on the positive strand of NCBI build 37. Linkage disequilibrium of the proxy SNPs evaluated in the eQTL study and the best *cis*-eQTL SNP in the region with the RA risk SNPs is indicated as r^2 values. When the expression probe was not assigned to any genes, the eQTL gene is labelled with a dash. **b**, *cis*-eQTL of T cells and monocytes in the RA risk SNPs. Significant *cis*-eQTLs of RA risk SNPs are indicated in bold (gene-based permutation $P < 0.05$).

* *cis*-eQTL of the proxy SNP (rs3807307, $r^2 = 0.96$) was evaluated.

† *cis*-eQTL of the proxy SNP (rs11557466, $r^2 = 0.98$) was evaluated.

Yoshida et al. Arthritis Research & Therapy 2014, 16:R75
http://arthritis-research.com/content/16/2/R75

arthritis research&therapy

RESEARCH ARTICLE Open Access

A *GC* polymorphism associated with serum 25-hydroxyvitamin D level is a risk factor for hip fracture in Japanese patients with rheumatoid arthritis: 10-year follow-up of the Institute of Rheumatology, Rheumatoid Arthritis cohort study

Shinji Yoshida[1,2†], Katsunori Ikari[1*†], Takefumi Furuya[1], Yoshiaki Toyama[2], Atsuo Taniguchi[1], Hisashi Yamanaka[1] and Shigeki Momohara[1]

Abstract

Introduction: Vitamin D deficiency has been reported to be common in patients with rheumatoid arthritis (RA) who have a higher prevalence of osteoporosis and hip fracture than healthy individuals. Genetic variants affecting serum 25-hydroxyvitamin D (25(OH)D) concentration, an indicator of vitamin D status, were recently identified by genome-wide association studies of Caucasian populations. The purpose of this study was to validate the association and to test whether the serum 25(OH)D-linked genetic variants were associated with the occurrence of hip fracture in Japanese RA patients.

Methods: DNA samples of 1,957 Japanese RA patients were obtained from the Institute of Rheumatology, Rheumatoid Arthritis (IORRA) cohort DNA collection. First, five single nucleotide polymorphisms (SNPs) that were reported to be associated with serum 25(OH)D concentration by genome-wide association studies were genotyped. The SNPs that showed a significant association with serum 25(OH)D level in the cross-sectional study were used in the longitudinal analysis of hip fracture risk. The genetic risk for hip fracture was determined by a multivariate Cox proportional hazards model in 1,957 patients with a maximum follow-up of 10 years (median, 8 years).

Results: Multivariate linear regression analyses showed that rs2282679 in *GC* (the gene encoding group-specific component (vitamin D binding protein)) locus was significantly associated with lower serum 25(OH)D concentration ($P = 8.1 \times 10^{-5}$). A Cox proportional hazards model indicated that rs2282679 in *GC* was significantly associated with the occurrence of hip fracture in a recessive model (hazard ratio (95% confidence interval) = 2.52 (1.05-6.05), $P = 0.039$).

Conclusions: A two-staged analysis demonstrated that rs2282679 in *GC* was associated with serum 25(OH)D concentration and could be a risk factor for hip fracture in Japanese RA patients.

* Correspondence: kikari@ior.twmu.ac.jp
†Equal contributors
[1]Institute of Rheumatology, Tokyo Women's Medical University, 10-22 Kawada, Shinjuku, Tokyo 162-0054, Japan
Full list of author information is available at the end of the article

Introduction

Vitamin D regulates calcium and phosphate homeostasis and reportedly has other roles in human physiology [1,2]. Vitamin D deficiency is associated with the occurrence of osteoporosis, autoimmune diseases, cardiovascular disease, type 1 and type 2 diabetes mellitus, and several types of cancer [3-7]. Vitamin D also plays an important role in the maintenance of the musculoskeletal system. It is positively associated with muscle strength and physical performance, and is negatively associated with fall and fracture risk [8-11].

Vitamin D deficiency has been reported to be common in patients with rheumatoid arthritis (RA), and more than 70% of Japanese patients with RA had vitamin D deficiency [12,13]. Significant associations of vitamin D deficiency were found with some independent clinical risk factors: female gender, younger age, high disability score in the Japanese version of the Health Assessment Questionnaire (J-HAQ), low serum total protein level, low serum total cholesterol level, high serum alkaline phosphate (ALP) level, and use of non-steroidal anti-inflammatory drugs (NSAIDs) [12].

Bone mineral density (BMD) is the major predictor of osteoporotic fracture, and previous studies have reported that patients with RA have a lower BMD and are at greater risk of hip fracture than healthy individuals [14,15]. We have previously shown that a high J-HAQ disability score, advanced age, history of total knee replacement (TKR), and low body mass index (BMI) were clinical risk factors for the occurrence of hip fracture in Japanese patients with RA [16].

Prior twin and family studies suggested that genetic factors also influence serum vitamin D concentration [17,18]. Genetic variants that affect serum 25-hydroxyvitamin D (25(OH)D) concentration, an indicator of vitamin D status, were recently identified in a meta-analysis of genome-wide association studies (GWAS) in Caucasian populations [19,20]. Though the presence of heterogeneity in genes related to RA has been suggested in many population-based studies [21], these associations remain unknown in the Japanese population.

The purpose of this study was to validate the possible association between genetic variants and serum 25(OH) D concentration and to test whether the serum 25(OH) D-linked variants were associated with the occurrence of hip fracture in Japanese patients with RA.

Methods

Study population

This study was a part of the Institute of Rheumatology, Rheumatoid Arthritis cohort study (IORRA), a single-institution-based, large-scale prospective observational cohort study with an enrollment of over 5,000 Japanese patients with RA, began in 2000 [12,16,22]. DNA samples of 1,957 Japanese RA patients were obtained from the IORRA DNA collection. All the patients satisfied the American College of Rheumatology 1987 revised criteria for RA. Tokyo Women's Medical University Genome Ethics Committee approved the present study and each individual signed an informed consent form after receiving a verbal explanation of the study.

SNP selection and genotyping

Five single nucleotide polymorphisms (SNPs) were selected from the results of recent genome-wide association studies (GWAS) that showed positive associations between serum 25(OH)D levels and the following: rs2282679 in the *GC* (group-specific component) locus, rs3829251, rs12785878 and rs1790349 in the *DHCR7/NADSYN1* (7-dehydrocholesterol reductase/nicotinamide-adenine dinucleotide synthetase 1) locus, and rs10741657 in the *CYP2R1* (cytochrome P450, family 2, subfamily R, polypeptide 1) locus [19,20]. Genotyping was performed by using the TaqMan fluorogenic 5′ nuclease assay according to the manufacturer's instructions (Applied Biosystems, Tokyo, Japan). Duplicate samples and negative controls were included to ensure accuracy of SNP genotyping. All polymerase chain reactions were performed by using the GeneAmp PCR System 9700 (Applied Biosystems), and endpoint fluorescent readings for TaqMan assays were done on an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems) as described elsewhere [22].

Measurement of serum 25(OH)D concentration

Serum 25(OH)D concentrations were measured in 899 of 1,957 patients from whom we obtained DNA samples. As vitamin D is synthesized in the skin under the influence of sunlight and the seasonal variance in serum levels of vitamin D is well known, serum samples were collected in the same season, spring of 2011 [23]. The DiaSorin 25(OH)D^{125}I radioimmunoassay kit was used for quantitative determination of serum 25(OH)D concentration.

Assessment of hip fracture

The occurrence of hip fracture after enrollment in IORRA was determined by the response to a patient questionnaire every six months from October 2000 to October 2010 with a maximum follow-up period of 10 years. The data were confirmed by review of medical records and radiographs as described elsewhere [16]. Only the occurrence of the first hip fracture reported by patients was included in this study. The occurrence of hip fracture caused by major trauma such as car accidents was excluded. A total of 39 hip fractures in 39 patients were identified and included in this study.

Statistical analyses

Stage 1: cross-sectional analyses of SNPs associated with serum 25(OH)D concentration

A two-staged analysis was used. First, cross-sectional associations were examined between serum 25(OH)D concentrations and risk alleles of each SNP using multivariate linear regression analyses (adjusted for independent non-genetic risk factors) in 899 of the 1,957 patients. The putative risk alleles were defined as the alleles that are associated with lower serum 25(OH)D concentration based on prior reports [19,20]. The number of risk alleles for genotyped SNPs (0, 1 and 2) was used to test the additive effect of the alleles on lower serum 25(OH)D concentration. The following factors that have been shown to be significantly associated with lower serum 25(OH)D levels were selected as independent non-genetic risk factors: gender, age, J-HAQ disability score, serum total protein level, serum total cholesterol level, serum ALP level, and NSAID use [12].

Stage 2: longitudinal association analyses of SNPs associated with occurrence of hip fracture

The SNP that showed a significant association with serum 25(OH)D concentration was selected to test the longitudinal association with the hip fracture event. The length of time from the date of enrollment in IORRA to the date of the occurrence of hip fracture was calculated. A multivariate Cox proportional hazards model adjusted for independent non-genetic risk factors that were associated with hip fracture was performed in the cohort of 1,957 patients [16]. The following factors at the time of hip fracture reported were used as the independent non-genetic risk factors: J-HAQ disability score, age, history of TKR, and BMI. The proportional hazards assumption for the Cox model was assessed using log-minus-log plots for survival analysis. All statistical analyses were performed using the R software package [24].

Results

Clinical and demographic characteristics of the patients

Demographic, clinical and therapeutic data describing the 1,957 patients at the time of enrollment in IORRA are shown in Table 1. In this cohort, serum 25(OH)D concentrations were measured in 899 patients in the spring of 2011 (Table 1). The median serum 25(OH)D concentration was 15.30 ng/mL (IQR, 12.10 to 20.10). During a maximum follow-up period of 10 years (median, 8.0 years; IQR, 4.5 to 10.0 years) 39 hip fractures in 1,957 patients were identified.

Genotyping

The overall genotyping success rate was 97.3% and the genotype concordance rate was 100% as assessed by duplicate samples. After the application of quality control criteria for genotyping (removed samples that consistently failed for >20% (1/5) of SNPs, removed SNPs with a call rate <95% after removing samples that consistently failed), 1,915 of the 1,957 samples and all polymorphisms were accepted for the analyses. The studied polymorphisms were found to be in Hardy-Weinberg equilibrium.

Stage 1: cross-sectional analyses of SNPs associated with serum 25(OH)D concentration

A multivariate linear regression analysis adjusted for independent non-genetic risk factors showed that the minor allele of rs2282679 (= C) in *GC* was significantly associated with lower serum 25(OH)D concentrations ($P = 8.1 \times 10^{-5}$, Table 2). The median serum 25(OH)D concentrations (ng/mL) for the genotypes of rs2282679 in *GC* were 16.1, 15.2 and 14.7, respectively for AA, AC and CC (Figure 1). The other SNPs did not show a significant association. The minor allele frequency of rs2282679 (*GC*) in RA patients (= 0.259) did not differ significantly from a Japanese control population (= 0.258, $P = 0.98$ by chi-squared test, n = 752), which was obtained from the DNA collection of the Pharma SNP Consortium, Tokyo, Japan, currently entrusted to the Health Science Research Resources Bank, Osaka, Japan, as described elsewhere [22].

Stage 2: longitudinal association analyses of SNPs associated with occurrence of hip fracture

A multivariate Cox proportional hazards regression model adjusted for the independent non-genetic risk factors indicated that homozygous for the risk allele of rs2282679 in the *GC* locus regarding low serum 25(OH)D concentration was significantly associated with the occurrence of hip fracture (hazard ratio (95% CI) = 2.52 (1.05, 6.05), $P = 0.039$) (Table 3, Figure 2). The association was still significant when the use of active vitamin D or bisphosphonate were included in the analysis as independent variables. The proportion of patients treated with active vitamin D or bisphosphonate did not differ significantly between homozygous for the risk allele of rs2282679 (CC) and the others (data not shown). Supplemental results of the Cox regression analyses for other SNPs that were not associated with serum 25(OH)D concentration did not show a significant association with the occurrence of hip fracture (Table 3, Figure 2).

Discussion

In this study, an association between a polymorphism of rs2282679 in the *GC* locus and serum 25(OH)D concentration was validated in Japanese patients with RA. Minor alleles of rs2282679 had additive effects on decreasing serum 25(OH)D concentrations. In addition, rs2282679 was significantly associated with the occurrence of hip fracture in Japanese patients with RA. This

Table 1 Demographic, clinical and therapeutic data at the time of an enrollment in IORRA and at the time of measurement of serum 25(OH)D concentration

Factor	At the time of enrollment in IORRA (n = 1957)	At the time of measurement of serum 25(OH)D concentration (spring 2011, n = 899)
Age, years	57.5 (49.5 to 64.6)	64.3 (56.7 to 70.6)
Sex, female	1668 (85.2)	766 (87.7)
Duration of disease, years	7.0 (2.0 to 14.0)	16.0 (11.0 to 23.0)
BMI, kg/m^2	21.2 (19.4 to 23.3)	21.2 (19.2 to 23.2)
DAS28	4.2 (3.3 to 5.0)	3.0 (2.4 to 3.8)
J-HAQ	0.8 (0.2 to 1.4)	0.6 (0.1 to 0.8)
RF, positive	1531 (81.8)	698 (80.0)
History of smoking, ever	641 (34.5)	235 (28.4)
History of TKR, ever	85 (4.3)	178 (20.4)
DMARDs use, ever	1670 (85.3)	777 (89.0)
Methotrexate use, ever	792 (40.8)	638 (73.1)
Biologic use, ever	12 (0.6)	155 (17.8)
Corticosteroid use, ever	932 (47.6)	399 (45.7)
Bisphosphonate use, ever	63 (3.2)	280 (32.1)
Active vitamin D use, ever	61 (3.1)	112 (12.8)
Serum total protein level, g/dL	Data not available	7.3 (7.0 to 7.6)
Serum total cholesterol level, mg/dL	Data not available	209.0 (188.0 to 231.0)
Serum alkaline phosphate level, IU/L	Data not available	258.0 (213.0 to 321.2)
NSAIDs use, ever	Data not available	515 (59.0)

Data are presented as median (IQR) or n (%). IORRA, Institute of Rheumatology, Rheumatoid Arthritis cohort; BMI, body mass index; DAS28, disease activity score in 28 joints; J-HAQ, Japanese version of the Health Assessment Questionnaire; RF, rheumatoid factor; TKR, total knee replacement; DMARD, disease modifying antirheumatic drug; NSAID, non-steroidal anti-inflammatory drug.

is the first report that a SNP P7 in the *GC* locus was associated with the risk for hip fracture.

The *GC* gene encodes the group-specific component known as the vitamin D binding protein (DBP) that plays an important role in the vitamin D metabolic pathway [25]. Most circulating vitamin D metabolites are bound to DBP to be transported to target organs. In the previous candidate gene studies and the recent GWAS, some *GC* polymorphisms were associated with serum 25(OH)D concentration, and the strongest association was observed for rs2282679 [19,20,26]. Our results provide supportive evidence that serum 25(OH)D concentration might partly be affected by a polymorphism of rs2282679 or the other variants that are in tight linkage disequilibrium with rs2282679.

Vitamin D is an important factor in mineral metabolism, bone growth and maintenance of the skeleton [1]. In addition, 1,25$(OH)_2$D, one of the vitamin D metabolites, has direct action on muscle strength and function modulated by vitamin D receptors expressed in human muscle tissue [8]. Therefore, vitamin D deficiency can lead to low bone density and muscle weakness, resulting in falls and fractures [8,27,28]. In many studies, vitamin D supplementation has been reported to reduce the risk for falls and fractures among older individuals [29,30].

To date, many genetic factors associated with low-trauma fracture including hip fracture have been reported in genome-wide meta-analysis studies [31,32]. Many variants with small effects may contribute to fracture risk, but only a few vitamin D-related genetic polymorphisms

Table 2 Multivariate linear regression analyses of each SNP associated with serum 25(OH)D concentration

Locus	SNP	MAF[†]	β	SE	*P*-value
GC	rs2282679	0.259 (A/C)	−0.13	0.033	8.1×10^{-5}
DHCR7/NADSYN1	rs3829251	0.375 (G/A)	−0.0031	0.032	0.92
	rs12785878	0.328 (G/T)	0.0057	0.032	0.86
	rs1790349	0.361 (A/G)	−0.016	0.032	0.63
CYP2R1	rs10741657	0.387 (G/A)	0.035	0.034	0.30

All analyses were adjusted for independent non-genetic risk factors: gender, age, Japanese version of the Health Assessment Questionnaire disability score, total protein level, total cholesterol level, alkaline phosphate level, and non-steroidal anti-inflammatory drugs use [12]. [†]Alleles are listed as major allele/minor allele. SNP, single nucleotide polymorphism; MAF, minor allele frequency; SE, standard error; *GC*, group-specific component; *DHCR7/NADSYN1*, 7-dehydrocholesterol reductase/nicotinamide-adenine dinucleotide synthetase 1; *CYP2R1*, cytochrome P450, family 2, subfamily R, polypeptide 1.

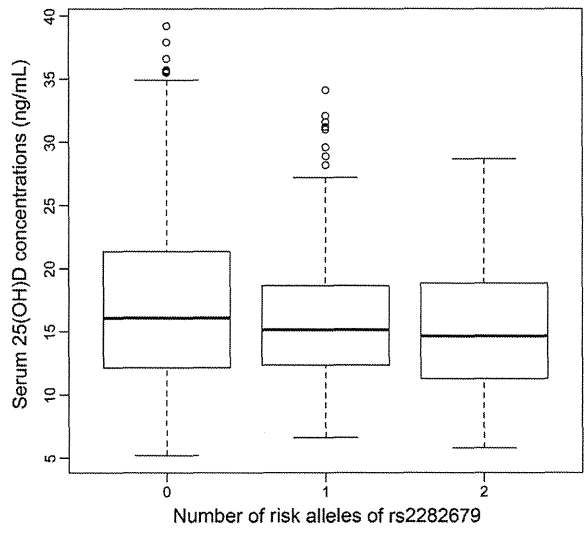


Figure 1 Boxplots representing the distribution of serum 25 (OH)D concentration according to the number of the risk allele of rs2282679 (minor allele, C) in the *GC* locus. Each box represents the IQR range of values, with the bold line showing the median value. The vertical lines show maximum and minimum values that fall within 1.5 box lengths, the open circles show extreme values >1.5 box plot lengths.

have been reported to be associated with fracture risk [33]. We explored the genetic risk of hip fracture in variants demonstrated to be associated with lower serum vitamin D concentration and found an association between rs2282679 in *GC* and the occurrence of hip fracture. Our results indicated that the risk allele carriers of the *GC* gene polymorphism tend to have low vitamin D levels that lead to greater risk of hip fracture.

Though *DHCR7/NADSYN1* and *CYP2R1* polymorphisms were associated with serum 25(OH)D concentration in the recent GWAS, we could not validate such an association in this study [19,20]. There are a number of possible explanations for the lack of an association. One is the insufficient statistical power to validate the associations. The number of samples was smaller than in the previous reports [20]. The SNP with highest statistical power to validate the association in this study was rs2282679 in the *GC* locus with a value of 0.72, and the others had comparatively lower statistical power (for example, 0.22 with rs3829251 in the *DHCR7/NADSYN1* locus). Another reason for the lack of an association is that all studied subjects were RA patients, whereas the recent GWAS were in healthy individuals [19,20]. Vitamin D is related to immunological processes, and vitamin D status has been reported to be associated with the risk of developing autoimmune diseases including RA [1,14]. In addition, serum vitamin D concentration has been shown to be lower in patients with greater disease activity [34]. Although the disease activity of the patients might affect the results of this study, DAS28 was not associated with serum 25(OH)D level in the studied population (data not shown). The difference in the genetic background between Caucasian and Japanese populations might also affect the results, which suggests genetic heterogeneity in *NADSYN1*, *DHCR7* and *CYP2R1*.

Table 3 Hip fracture risk for rs2282679 (GC) and the other SNPs adjusted for independent non-genetic risk factors

Locus	SNP	HR	95% CI	*P*-value
GC	rs2282679	2.52	1.05, 6.05	0.039
DHCR7/NADSYN1	rs3829251	1.00	0.39, 2.56	1.00
	rs12785878	0.71	0.22, 2.30	0.56
	rs1790349	1.04	0.40, 2.65	0.94
CYP2R1	rs10741657	1.31	0.58, 3.02	0.51

All analyses were adjusted for independent non-genetic factors: Japanese version of the Health Assessment Questionnaire disability score, age, history of total knee replacement, and body mass index [16]. SNP, single nucleotide polymorphism; HR, hazard ratio; *GC*, group-specific component; *DHCR7/NADSYN1*, 7-dehydrocholesterol reductase/nicotinamide-adenine dinucleotide synthetase 1; *CYP2R1*, cytochrome P450, family 2, subfamily R, polypeptide 1.

The strength of this study is that the datasets were relatively large and based on a single-institution cohort study of Japanese patients with RA. Serum 25(OH)D concentration was measured in the same season of the same year. Therefore, the differences between regions, heterogeneity and seasons had less influence on the results.

The potential limitation of this study is that the serum 25(OH)D concentration data were available from only 899 of the 1,957 patients with DNA samples, and the study on serum 25(OH)D concentration was a cross-sectional study. The smaller sample size reduced the statistical power to detect minor effects on events. Though the measurement from multiple time points would provide more valid estimates of the results, there was only one blood sample assayed for serum 25 (OH)D concentration for each person. Further studies are required to confirm these associations.

Conclusion

In conclusion, our data demonstrated that rs2282679 in *GC* was associated with both serum 25(OH)D concentration and the occurrence of hip fracture in Japanese patients with RA. These results might contribute to a better understanding of the biological impact of genetic variation within the vitamin D metabolic pathway.