Genetics of rheumatoid arthritis contributes to biology and drug discovery

Citation

Published Version

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:12785839

Terms of Use

Share Your Story

Accessibility

Genetics of rheumatoid arthritis contributes to biology and drug discovery

A full list of authors and affiliations appears at the end of the article.

Abstract

A major challenge in human genetics is to devise a systematic strategy to integrate disease-associated variants with diverse genomic and biological datasets to provide insight into disease pathogenesis and guide drug discovery for complex traits such as rheumatoid arthritis (RA)[1].

[*]Correspondence authors: Robert M. Plenge and Yukinori Okada, Robert M. Plenge, MD, PhD, *Director, Genetics & Genomics*, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Suite 168, Boston, MA 02115, USA., Telephone: +1-617-525-4451, Fax: +1-617-525-4488, robert.plenge@merck.com, Yukinori Okada, MD, PhD, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Suite 255, Boston, MA 02115, USA., Telephone: +1-617-525-4055, Fax: +1-617-525-4488, yokada@broadinstitute.org.
[73]Full list of contributing authors for the RACI and GARNET consortia is provided in **Supplementary Information**.

Supplementary Information is available in the online version of the paper at www.nature.com/nature.

Author contributions

Web resources

Summary statistics from the GWAS meta-analysis, source codes, and data sources, 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, https://slegen.phs.wfubmc.edu/public/index.cfm

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/

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

Here, we performed a genome-wide association study (GWAS) meta-analysis in a total of >100,000 subjects of European and Asian ancestries (29,880 RA cases and 73,758 controls), by evaluating ~10 million single nucleotide polymorphisms (SNPs). We discovered 42 novel RA risk loci at a genome-wide level of significance, bringing the total to 101[2–4]. We devised an *in-silico* pipeline using established bioinformatics methods based on functional annotation[5], cis-acting expression quantitative trait loci (cis-eQTL)[6], and pathway analyses[7–9] – as well as novel methods based on genetic overlap with human primary immunodeficiency (PID), hematological cancer somatic mutations and knock-out mouse phenotypes – to identify 98 biological candidate genes at these 101 risk loci. We demonstrate that these genes are the targets of approved therapies for RA, and further suggest that drugs approved for other indications may be repurposed for the treatment of RA. Together, this comprehensive genetic study sheds light on fundamental genes, pathways and cell types that contribute to RA pathogenesis, and provides empirical evidence that the genetics of RA can provide important information for drug discovery.

We conducted a three-stage trans-ethnic meta-analysis (Extended Data Fig. 1). Based on the polygenic architecture of RA[10] and shared genetic risk among different ancestry[3,4], we hypothesized that combining GWAS of European and Asian ancestry would increase power to detect novel risk loci. In Stage I, we combined 22 GWAS for 19,234 cases and 61,565 controls of European and Asian ancestry[2–4]. We performed trans-ethnic, European-specific, and Asian-specific GWAS meta-analysis by evaluating ~10 million SNPs[11]. Characteristics of the cohorts, genotyping platforms, quality control (QC) criteria are described in Extended Data Table 1 (overall $\lambda_{GC} < 1.075$).

Stage I meta-analysis identified 57 loci that satisfied a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$, including 17 novel loci (Extended Data Fig. 2). We then conducted a two-step replication study (Stage II for *in-silico* and Stage III for *de-novo*) in 10,646 RA cases and 12,193 controls for the loci with $P < 5.0 \times 10^{-6}$ in Stage I. In a combined analysis of Stages I–III, we identified 42 novel loci with $P < 5.0 \times 10^{-8}$ in either of the trans-ethnic, European, or Asian meta-analyses. This increases the total number of RA risk loci to 101 (Table 1 and Supplementary Table 1).

Comparison of 101 RA risk loci revealed significant correlations of risk allele frequencies (RAF) and odds ratios (OR) between Europeans and Asians (Extended Data Fig. 3a–c; Spearman's $\rho = 0.67$ for RAF and 0.76 for OR; $P < 1.0 \times 10^{-13}$), although 5 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 95% confidence intervals [95%CI] of OR). In the population-specific genetic risk model, the 100 RA risk loci outside of the major histocompatibility complex (MHC) region[12] explained 5.5% and 4.7% of heritability in Europeans and Asians, respectively, with 1.6% of the heritability by the novel loci. The trans-ethnic genetic risk model, based on RAF from one population but OR from the other population, could explain the majority (>80%) of the known heritability in each population (4.7% for Europeans and 3.8% for Asians). These observations support our hypothesis that the genetic risk of RA is shared, in general, among Asians and Europeans

We assessed enrichment of 100 non-MHC RA risk loci in epigenetic chromatin marks (Extended Data Fig. 3d)[13]. Of 34 cell types investigated, we observed significant enrichment of RA risk alleles with trimethylation of histone H3 at lysine 4 (H3K4me3) peaks in primary CD4$^+$ regulatory T cells (T$_{reg}$ cells; $P < 1.0 \times 10^{-5}$). For the RA risk loci enriched with T$_{reg}$ H3K4me3 peaks, we incorporated the epigenetic annotations along with trans-ethnic differences in patterns of linkage disequilibrium (LD) to fine-map putative causal risk alleles (Extended Data Fig. 3e–f).

We found that approximately two-thirds of RA risk loci demonstrated pleiotropy with other human phenotypes (Extended Data Fig. 4), including immune-related diseases (e.g., vitiligo, primary biliary cirrhosis), inflammation-related or hematological biomarkers (e.g., fibrinogen, neutrophil counts) and other complex traits (e.g., cardiovascular diseases).

Each of 100 non-MHC RA risk loci contains on average ~4 genes in the region of LD (in total 377 genes). To systematically prioritize the most likely biological candidate gene, we devised an *in silico* bioinformatics pipeline. In addition to the published methods that integrate data across associated loci[7,8], we evaluated several biological datasets to test for enrichment of RA risk genes, which help to pinpoint a specific gene in each loci (Extended Data Fig. 5–6, Supplementary Tables 2–4).

We firstly conducted functional annotation of RA risk SNPs. Sixteen percent of SNPs were in LD with missense SNPs ($r2 > 0.80$; Extended Data Fig. 5a–b). The proportion of missense RA risk SNPs was higher compared to a set of genome-wide common SNPs (8.0%), and relatively much higher in the explained heritability (~26.8%). Using cis-eQTL data obtained from peripheral blood mononuclear cells (5,311 individuals)[6] and from CD4$^+$ T cells and CD14$^+$16$^-$ monocytes (212 individuals), we found that RA risk SNPs in 44 loci showed cis-eQTL effects (FDR-q or permutation $P < 0.05$; Extended Data Table 2).

Second, we evaluated whether genes from RA risk loci overlapped with PID genes[14], and observed significant overlap ($14/194 = 7.2\%$, $P = 1.2\times10^{-4}$; Fig. 1a and Extended Data Fig. 5c). Classification categories of PID genes showed different patterns of overlap: highest proportion of overlap in "immune dysregulation" ($4/21 = 19.0\%$, $P = 0.0033$) but no overlap in "innate immunity".

Third, we evaluated overlap with cancer somatic mutation genes[15], under the hypothesis that genes with cell growth advantages may contribute to RA development. Among 444 genes with registered cancer somatic mutations[15], we observed significant overlap with hematological cancers ($17/251 = 6.8\%$, $P = 1.2\times10^{-4}$; Fig. 1b and Extended Data Fig. 5d), but not from non-hematological cancers ($6/221 = 2.7\%$, $P = 0.56$).

Fourth, we evaluated overlap with genes implicated in knockout mouse phenotypes[16]. Among the 30 categories of phenotypes[16], we observed 3 categories significantly enriched with RA risk genes ($P < 0.05/30 = 0.0017$): "hematopoietic system phenotype", "immune system phenotype", and "cellular phenotype" (Extended Data Fig. 5e).

Fifth, we conducted molecular pathway enrichment analysis (Fig. 1c and Extended Data Fig. 5f). We observed enrichment (FDR-q < 0.05) for T cell related pathways, consistent with cell-specific epigenetic marks, as well as enrichment for B cell and cytokine signaling pathways (e.g., IL-10, interferon, GM-CSF). For comparison, our previous RA GWAS meta-analysis[2] did not identify the B cell and cytokine signaling pathways, thereby indicating that as more loci are discovered, further biological pathways are identified.

Based on these new findings, we adopted the following 8 criteria to prioritize each of the 377 genes from the 100 non-MHC RA risk loci (Fig. 2 and Extended Data Fig. 6a–c): (i) genes with RA risk missense variant ($n = 19$); (ii) cis-eQTL genes ($n = 51$); (iii) genes prioritized by PubMed text-mining[7] ($n = 90$); (iv) genes prioritized by protein-protein interaction (PPI)[8] ($n = 63$); (v) PID genes ($n = 15$); (vi) hematological cancer somatic mutation genes ($n = 17$); (vii) genes prioritized by associated knockout mouse phenotypes ($n = 86$); and (viii) genes prioritized by molecular pathway analysis[9] ($n = 35$).

Ninety-eight genes (26.0%) had a score 2, which we defined as "candidate biological RA risk genes". Nineteen loci included multiple biological RA risk genes (e.g., *IL3* and *CSF2* at 5q31), while no biological gene was selected from 40 loci (Supplementary Table. 5).

To provide empirical evidence of the pipeline, we evaluated relationships of the gene scores to independent genomic or epigenetic information. Genes with higher biological scores were more likely to be the nearest gene to the risk SNP (18.6% for gene score <2 and 49.0% for gene score 2; $P = 2.1 \times 10^{-8}$), and also to be included in the region where RA risk SNPs were overlapping with H3K4me3 T_{reg} peaks (41.9% for gene score <2 and 57.1% for gene score 2; $P = 0.034$). Further, T_{reg} cells demonstrated the largest increase in overlapping proportions with H3K4me3 peaks for increase of biological gene scores compared to other cell types (Extended Data Fig. 6d).

Finally, we evaluated the potential role of RA genetics in drug discovery. We hypothesized that if human genetics is useful for drug target validation, then it should identify existing approved drugs for RA. To test this "therapeutic hypothesis"[1], we obtained 871 drug target genes corresponding to approved, clinical trial, or experimental drugs of human diseases (Supplementary Table 6)[17,18]. We evaluated whether any of the protein-products from the biological RA risk genes, or any genes from a direct PPI network with them (Fig. 3a), are the pharmacologically active targets of approved RA drugs (Extended Data Fig. 7a).

Twenty-seven drug target genes of approved RA drugs demonstrated significant overlap with 98 biological RA risk genes and 2,332 genes from the expanded PPI network (18 genes overlapped; 3.7-fold enrichment by permutation analysis, $P < 1.0 \times 10^{-5}$; Fig. 3b). For comparison, all drug target genes (regardless of disease indication) overlapped with 247 genes, which is 1.7-fold enrichment more than expected by chance, but less than 2.2-fold enriched compared to overlap of the target genes of RA drugs ($P = 0.0035$). Examples of approved RA therapies identified by this analysis include tocilizumab[19,20] (anti-IL6R), tofacitinib[21] (JAK3 inhibitor), and abatacept[21] (CTLA4-Ig; Fig. 3c and Extended Data Fig. 8).

We also assessed how approved drugs for other diseases might be connected to biological RA risk genes. We highlight *CDK6* and *CDK4*, targets of three approved drugs for different types of cancer[22] (Fig. 3d). In support for repurposing, one *CDK6/CDK4* inhibitor, flavopiridol, has been shown to ameliorate disease activity in animal models of RA[22]. Further, the biology is plausible, since several approved RA drugs were initially developed for cancer treatment and then repurposed for RA (e.g., rituximab). While further investigations are necessary, we propose that target genes/drugs selected by this approach could represent promising candidates for novel drug discovery for RA treatment.

We note that a non-random distribution of drug-to-disease indications in the databases could potentially bias our results. Namely, since RA risk genes are enriched for genes with immune function, spurious enrichment with drug targets could occur if the majority of drug indications in databases were for immune-mediated diseases or immune-related target genes. However, such enrichment was not evident in our analysis (~11% for drug indications and ~9% for target genes; Extended Data Fig. 7b).

In summary, through a comprehensive genetic study with >100,000 subjects, we identified 42 novel RA risk loci and provided novel insight into RA pathogenesis. We particularly highlight the role of genetics for drug discovery. While there are anecdotal examples[1,23], our study provides a systematic approach by which human genetic data can be efficiently integrated with other biological information to derive biological insights and drug discovery.

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% for 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 as RA by a professional rheumatologist. The 19,234 RA cases and 61,565 controls enrolled in the Stage I trans-ethnic GWAS meta-analysis were obtained from 22 studies from 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, 3 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 6 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 II *in silico* replication study were obtained from 2 studies from Europeans (2,780 RA cases and 4,700 controls from Genentech and SLEGEN) and Asians (928 RA cases and 835 controls from China) [manuscript submitted]. The 6,938 RA cases and 6,658 controls enrolled in the Stage III *de novo* replication study were obtained from 2 studies from 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 I, Stage II and Stage III studies were confirmed to be independent through analysis of overlapping SNP markers. Any duplicate subjects were removed from the Stage II and Stage III replication studies, leading to slightly different sample sizes compared to 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 at Extended Data Fig. 1 and Extended Data Table 1a, as well as the previous reports[2–4].

Genotyping, quality control, and genotype imputation of GWAS data

Genotyping platforms and QC criteria of GWAS, including cutoff values for sample call rate, SNP call rate, 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 analyzed based on the same analytical protocol, including exclusion of closely-related subjects and outliers in terms of ancestries, as described elsewhere[3]. After applying QC criteria, whole-genome genotype imputation was performed using 1000 Genome 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 kbp of duplicated regions between neighboring chunks. Immunochip data were split into ~2,000 chunks which 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 X-chromosome were imputed for males and females separately. We excluded imputed SNPs which were duplicated between chunks, SNPs with minor allele frequency (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 in 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 the European ancestry[11].

Stage I trans-ethnic genome-wide meta-analysis

Associations of the 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 v1.0.16 (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 analyzed 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 (Beta) and the standard errors (SE) 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 Mbp of known RA loci or in the MHC region (chromosome 6, 25–35 Mbp). While 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 (λ_{GC} < 1.075 and λ_{GC_1000} < 1.005; Extended Data Table 1b).

As for the definition of known RA risk loci in this study, we included the loci which showed significant associations in one of the previous studies ($P < 5.0 \times 10^{-8}$) or which 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*[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 I 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 II and Stage III replication studies

In silico (Stage II) and *de novo* (Stage III) 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 I trans-ethnic, European, or Asian GWAS meta-analysis were selected for the Stage II *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 datasets, a proxy SNP with the highest LD ($r2 > 0.80$) was alternatively assessed. GWAS QC, genotype imputation, association analysis were assessed in the same manner as in the Stage I GWAS. For the 60 loci that demonstrated suggestive associations in the combined result of Stage I GWAS meta-analysis and the Stage II *in silico* replication study but not included as a known RA risk locus, we calculated statistical power to newly achieve genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ for Europeans and Asians separately, which were estimated based on the allele frequencies, OR, and *de novo* replication sample sizes of the populations. We then selected top 20 SNPs with the highest

statistical power for Europeans and Asians separately (in total 32 SNPs), and conducted the Stage III *de novo* replication study. Genotyping methods, quality control, and confirmation of subject independence in the Stage III *de novo* replication study were described in the previous studies[2,3]. The combined study of the Stage I GWAS meta-analysis and the Stage II and III 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), OR, and explained heritability were conducted using the results of the Stage I GWAS meta-analysis of Europeans and Asians. Correlations of RAF and OR were evaluated using Spearman's correlation test. OR were defined based on minor alleles in Europeans. Explained heritability was estimated by applying a liability-threshold model assuming disease prevalence of 0.5%[10] and using RAF and OR of the population(s) according to the genetic model. For population-specific genetic risk model, RAF and OR of the same population was used. For trans-ethnic genetic risk model, RAF of the population but 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 LD ($r2 > 0.80$) with them were enriched in overlap with H3K4me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) assay peaks of 34 cell types obtained from NIH 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 I GWAS meta-analysis of both Europeans and Asians with same directional effects of alleles (outside of the MHC region). As for fine-mapping using LD structure differences between the populations, we calculated average numbers of the SNPs in LD ($r2 > 0.80$) in Europeans, in Asians, in both Europeans and Asians, separately.

As for fine-mapping using H3K4me3 peaks of T_{reg} primary cells, we firstly evaluated H3K4me3 peak overlap enrichment of the SNPs in LD (in Europeans and Asians) compared to the neighboring SNPs (± 2 Mbp). We fixed the SNP positions but physically slid H3K4me3 peak positions by 1 kbp bins within ± 2 Mbp regions of the risk SNPs, and calculated overlap of the SNPs in LD with H3K4me3 peaks for each sliding step, and evaluated the significance of overlap in the original peak positions by 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 which were in LD in both Europeans and Asians and also included in H3K4me3 peaks.

Pleiotropy analysis

We downloaded phenotype-associated SNPs and phenotype information from NHGRI GWAS catalogue database[26] on January 31, 2013. We selected significantly associated 4,676 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 the single phenotype (eg. from "Urate levels", "Uric acid levels", and "Renal function-related traits (urea)" into "Urate levels"), or splitting the merged phenotypes into the sub-categorical phenotypes (eg. 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 SNPs and the RA risk SNPs identified by our study (located outside of the MHC region), we defined the genetic region based on ±25 kbp of the SNP or the neighboring SNP positions in moderate LD with it in Europeans or Asians ($r2 > 0.50$). If multiple different SNPs with overlapping regions were registered for the same NHGRI phenotype, they were merged into the single region. We defined "region-based pleiotropy" if two phenotype-associated SNPs shared part of their genetic regions or shared any UCSC ref gene(s) (hg19) partly overlapping with each of the regions (Extended Data Fig. 4a). We defined "allele-based pleiotropy" if two phenotype-associated SNPs were in LD in Europeans or Asians ($r2 > 0.80$). We defined the direction of effect as "concordant" with RA risk if the RA risk allele also leads to increased risk of the NHGRI 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 NHGRI 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 number of the observed region-based pleiotropy with RA for each of the iteration steps. For null distribution estimation, we did not include the SNPs associated with several autoimmune diseases which were previously reported to share pleiotropic associations with RA (Crohn's disease, type 1 diabetes, multiple sclerosis, celiac disease, systemic lupus erythematosus, ulcerative colitis, and psoriasis)[2].

Prioritization of biological candidate genes from RA risk loci

For the RA risk SNPs outside of the MHC region, functional annotations were conducted by Annovar (hg19). The RA risk SNPs were classified if any of the SNPs in LD ($r2 > 0.80$) 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 randomly selected 10,000 genome-wide common SNPs (MAF > 0.05 in Europeans or Asians).

We then assessed cis-eQTL effects by referring two eQTL datasets; the study for PBMC obtained from 5,311 European subjects[6] and newly-generated cell-specific eQTL analysis for $CD4^+$ T cells and $CD14^+16^-$ monocytes from 212 European subjects (ImmVar project; Raj *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 LD with the highest $r2$ 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 defined by the International Union of Immunological Societies (IUIS) 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 January 31, 2013 (Supplementary Table 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 kbp or $r2 > 0.50$), and evaluated overlap with PID categories, cancer phenotypes with registered somatic mutations, and phenotype labels of knockout mouse genes with human orthologs. 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 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 I) and the previous meta-analysis of RA[2], separately. Since 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 newly imputed genotype data based on the 1000 Genome 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 8 selection criteria and calculating the number of the satisfied criteria: (i) genes of which RA risk SNPs or any of the SNPs in LD ($r2 > 0.80$) with them were annotated as missense variants; (ii) genes of which significant cis-eQTL of either of PBMC, T cell, and monocyte were observed for RA risk SNPs (FDR-q < 0.05 for PBMC and permutation $P < 0.05$ for T cells and monocytes); (iii) genes prioritized by PubMed text-mining using GRAIL[7] with gene-based $P < 0.05$; (iv) genes prioritized by PPI network using DAPPLE[8] with gene-based $P < 0.05$; (v) PID genes[14]; (vi) hematological cancer somatic mutation genes[15]; (vii) genes of which 2 of associated phenotype labels ("hematopoietic system phenotype", "immune system phenotype", and "cellular phenotype"; $P < 1.0 \times 10^{-4}$) were observed for knockout mouse[16]; and (viii) genes prioritized by molecular pathway analysis using MAGENTA[9] which were included in the significantly enriched pathways (FDR-q < 0.05) with gene-based $P < 0.05$. Since these criteria showed weak correlations with each other ($R2 < 0.26$; 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 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 LD with it) of the gene overlapped with H3K4me3 histone peaks of cell types. Difference of proportions to be the nearest gene between 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 (ver 2.15.2). Difference of proportions of the 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 Therapeutic Targets Database (TTD)[18] databases on January 31, 2013, as well as additional literature searches. We selected drug target genes which had pharmacological activities (for the genes from DrugBank) and human orthologs, and was 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 based on discussions with professional rheumatologists (Extended Data Fig. 7a). We extracted genes in direct PPI with biological RA risk genes by using InWeb database[27]. In order to take account of potential dependence between PPI genes and drug target genes, overlap of biological RA risk genes and genes in

Affiliations

[1]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [4]Department of Statistics, Harvard University, Cambridge, MA, USA. [5]Centre for Cancer Research, Monash Institute of Medical Research, Monash University, Clayton, Victoria, Australia. [6]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA. [7]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [8]Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [9]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan. [10]Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [11]Immunology Biomarkers Group, Genentech, South San Francisco, CA, USA. [12]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA. [13]Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. [14]New York University Hospital for Joint Diseases, New York, NY, USA. [15]Department of Medicine, Albany Medical Center and The Center for Rheumatology, Albany, NY, USA. [16]Division of Rheumatology, Department of Medicine, New York, Presbyterian Hospital, College of Physicians and Surgeons, Columbia University, New York, NY, USA. [17]Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai 200003, China. [18]Department of Pharmacology, Second Military Medical University, Shanghai, 200433, China [19]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia. [20]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. [21]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. [22]Toronto General Research Institute, Toronto, Canada. [23]Department of Medicine, University of Toronto, Toronto, Canada. [24]Department of Medicine, Mount Sinai Hospital and University of Toronto, Toronto, Canada. [25]Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen, the Netherlands. [26]Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia. [27]Division of Endocrinology, Children's Hospital, Boston, MA, USA. [28]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China. [29]The Department of Psychiatry at Mount Sinai School of Medicine, New York, USA. [30]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [31]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [32]Department of Rheumatology and Clinical Immunology, Arthritis Center Twente, University Twente & Medisch Spectrum Twente, Enschede, The Netherlands. [33]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, The Netherlands. [34]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands. [35]Service de Rhumatologie et INSERM U699 Hôpital Bichat Claude Bernard, Assistance Publique des Hôpitaux de Paris, Paris, France. [36]Université Paris 7-Diderot, Paris, France. [37]Université Paris-Sud, Orsay, France. [38]APHP–Hôpital Bicêtre, INSERM U1012, Le Kremlin Bicêtre, Paris, France. [39]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [40]AMC/ University of Amsterdam, Amsterdam, the Netherlands. [41]GlaxoSmithKline, Stevenage, U.K. [42]University of Cambridge, Cambridge, U.K. [43]Institut National de

la Santé et de la Recherche Médicale (INSERM) U1012. [44]Université Paris-Sud 11, Rhumatologie, Hôpital Bicêtre, Assistance Publique-Hôpitaux de Paris (AP-HP), Le Kremlin Bicêtre, France. [45]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea. [46]Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada, Spain. [47]Department of Rheumatology, Hospital Marques de Valdecilla, IFIMAV, Santander, Spain. [48]Hospital Clinico San Carlos, Madrid, Spain. [49]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [50]Department of Rheumatology, Umeå University, Umeå, Sweden. [51]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [52]Section of Rheumatology, Boston University School of Medicine, Boston, MA, USA. [53]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, MA, USA. [54]Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. [55]Université Paris 13 Sorbonne Paris Cité, UREN (Nutritional Epidemiology Research Unit), Inserm (U557), Inra (U1125), Cnam, Bobigny, France. [56]McGill University and Génome Québec Innovation Centre, Montréal, Canada. [57]Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [58]National Institute for Health Research, Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. [59]Department of Clinical Immunology and Rheumatology & Department of Genome Analysis, Academic Medical Center/University of Amsterdam, Amsterdam, The Netherlands. [60]Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, USA. [61]Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [62]Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto, Japan. [63]Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [64]Rheumatology Unit, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden. [65]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, NY, USA. [66]NIHR Manchester Musculoskeletal Biomedical, Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. [67]Section of Genetic Medicine, University of Chicago, Chicago, IL, USA. [68]Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL, USA. [69]Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan. [70]Core Research for Evolutional Science and Technology (CREST) program, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan. [71]Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan. [72]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan.
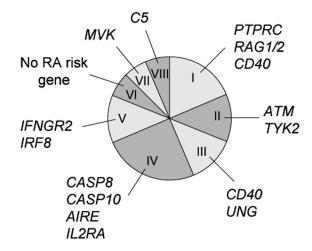
Acknowledgments

References

1. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013; 12:581–594. [PubMed: 23868113]

2. Stahl EA, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010; 42:508–514. [PubMed: 20453842]

3. Okada Y, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet. 2012; 44:511–516. [PubMed: 22446963]

4. Eyre S, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet. 2012; 44:1336–1340. [PubMed: 23143596]

5. Ferreira RC, et al. Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and Influences Risk of Diverse Inflammatory Diseases. PLoS Genet. 2013; 9:e1003444. [PubMed: 23593036]

6. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45:1238–1243. [PubMed: 24013639]

7. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]

8. Rossin EJ, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]

9. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. 2010; 6:e1001058. [PubMed: 20714348]

10. Stahl EA, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet. 2012; 44:483–489. [PubMed: 22446960]

11. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

12. Raychaudhuri S, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012; 44:291–296. [PubMed: 22286218]

13. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet. 2013; 45:124–130. [PubMed: 23263488]

14. Parvaneh N, Casanova JL, Notarangelo LD, Conley ME. Primary immunodeficiencies: a rapidly evolving story. J Allergy Clin Immunol. 2013; 131:314–323. [PubMed: 23374262]

15. Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–D950. [PubMed: 20952405]

16. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. 2012; 40:D881–D886. [PubMed: 22075990]

17. Knox C, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2011; 39:D1035–D1041. [PubMed: 21059682]

18. Zhu F, et al. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res. 2012; 40:D1128–D1136. [PubMed: 21948793]

19. Smolen JS, et al. Consensus statement on blocking the effects of interleukin-6 and in particular by interleukin-6 receptor inhibition in rheumatoid arthritis and other inflammatory conditions. Ann Rheum Dis. 2013; 72:482–492. [PubMed: 23172750]

20. Nishimoto N, et al. Study of active controlled tocilizumab monotherapy for rheumatoid arthritis patients with an inadequate response to methotrexate (SATORI): significant reduction in disease activity and serum vascular endothelial growth factor by IL-6 receptor inhibition therapy. Mod Rheumatol. 2009; 19:12–19. [PubMed: 18979150]

21. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011; 365:2205–2219. [PubMed: 22150039]

22. Sekine C, et al. Successful treatment of animal models of rheumatoid arthritis with small-molecule cyclin-dependent kinase inhibitors. J Immunol. 2008; 180:1954–1961. [PubMed: 18209094]

23. Sanseau P, et al. Use of genome-wide association studies for drug repositioning. Nat Biotechnol. 2012; 30:317–320. [PubMed: 22491277]

24. Arnett FC, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988; 31:315–324. [PubMed: 3358796]

25. Okada Y, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. Nat Genet. 2012; 44:904–909. [PubMed: 22797727]

26. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–9367. [PubMed: 19474294]

27. Lage K, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol. 2007; 25:309–316. [PubMed: 17344885]

28. Ueda H, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature. 2003; 423:506–511. [PubMed: 12724780]

29. Elliott P, et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA. 2009; 302:37–48. [PubMed: 19567438]

30. Cortes A, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 2013; 45:730–738. [PubMed: 23749187]
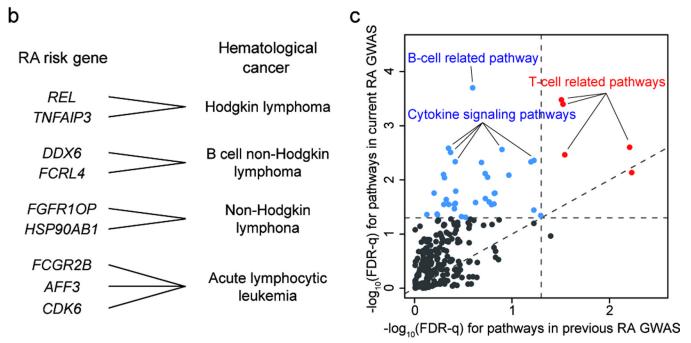
Figure 1. Overlap of RA risk loci with PID, hematological cancer somatic mutation, and molecular pathways

a, Overlap of RA risk genes with PID genes, subset by PID categories (I-VIII). **b**, Examples of overlap of hematological cancer somatic mutation genes with RA risk genes. **c**, Comparisons of molecular pathway analysis results between the current trans-ethnic meta-analysis (y-axis) and the previous meta-analysis for rheumatoid arthritis (x-axis)[2]. Each dot represents a molecular pathway. Dotted line represents FDR-q = 0.05 or $y = x$.

RA risk SNP (cytoband)	Gene	Score
chr1:2523811 (1p36)	TNFRSF14	4
rs2301888 (1p36)	PADI4	2
rs2476601 (1p13)	PTPN22	5
rs2228145 (1q21)	IL6R	5
chr1:161644258 (1q23)	FCGR2B	5
rs17668708 (1q31)	PTPRC	6
rs34695944 (2p16-p15)	REL	4
rs9653442 (2q11)	AFF3	4
rs11889341 (2q32)	STAT4	3
rs6715284 (2q33)	CFLAR	3
rs1980422 (2q33)	CD28	4
rs3087243 (2q33)	CTLA4	4
rs45475795 (4q26-q27)	IL2	5
rs657075 (5q31)	IL3	4
rs657075 (5q31)	CSF2	4
rs2233424 (6p21)	NFKBIE	4
rs7752903 (6q23)	TNFAIP3	6
rs1571878 (6q27)	CCR6	2
rs4272 (7q21)	CDK6	4
chr7:128580042 (7q32)	IRF5	4
rs10985070 (9q33)	TRAF1	4
rs10985070 (9q33)	C5	4
rs706778 (10p15)	IL2RA	5
rs331463 (11p12)	TRAF6	4
rs331463 (11p12)	RAG1	4
rs508970 (11q12)	CD5	4
chr11:107967350 (11q22)	ATM	4
rs773125 (12q13)	CDK2	3
rs1633360 (12q13-q14)	CDK4	3
rs10774624 (12q24)	SH2B3	5
chr17:38031857 (17q12-q21)	IKZF3	4
chr17:38031857 (17q12-q21)	CSF3	4
rs8083786 (18p11)	PTPN2	3
rs34536443 (19p13)	ICAM1	4
rs34536443 (19p13)	TYK2	6
rs4239702 (20q13)	CD40	6
rs73194058 (21q22)	IFNGR2	6
rs2236668 (21q22)	ICOSLG	5
rs2236668 (21q22)	AIRE	4
rs3218251 (22q12)	IL2RB	3
rs5987194 (Xq28)	IRAK1	3

Column headers — Biological gene criteria: RA risk missense variant, Cis-eQTL, PubMed text-mining, Protein-protein interaction, Primary immunodeficiency, Hematological cancer, Knockout mouse phenotype, Molecular pathway. Nearest gene from RA risk SNP. Overlap with H3K4me3 peaks: T_{reg} primary cells, CD4+ memory primary cells, CD4+ naïve primary cells, CD8+ memory primary cells, CD8+ naïve primary cells, CD34+ primary cells, CD34+ cultured cells, Mobilized CD34+ primary cells, CD19+ primary cells, CD3+ primary cells. Drug target gene, RA drug target gene, PPI with RA drug target gene.

Figure 2. Prioritized biological RA risk genes

Representative biological RA risk genes. We list the summary gene score derived from individual criterion (filled red box indicates criterion satisfied; 98 genes with score 2 out of 377 genes included in the RA risk loci were defined as "biological candidate genes"; see details in Extended Data Fig. 6). Filled blue box indicates the nearest gene to the RA risk SNP. Filled green boxes indicate overlap with H3K4me3 peaks in immune-related cells. Filled purple boxes indicate overlap with drug target genes. Full results are in Supplementary Table 5.
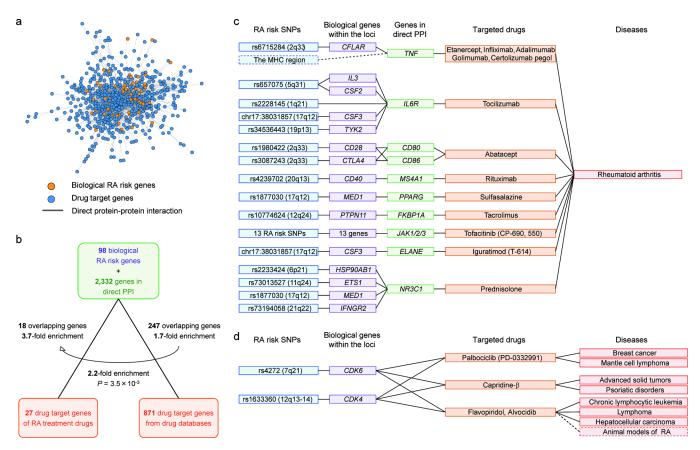
Figure 3. Connection of biological RA risk genes to drug targets
a, PPI network of biological RA risk genes and drug target genes. **b**, Overlap and relative enrichment of 98 biological RA risk genes with targets of approved RA drugs and with all drug target genes. Enrichment was more apparent than that from all 377 RA risk genes (Extended Data Fig. 7c). **c**, Connections between RA risk SNPs (blue), biological genes (purple), genes from PPI (green), and approved RA drugs (orange). Full results are in Extended Data Fig. 8. **d**, Connections between RA genes and drugs indicated for other diseases.

Table 1

Novel rheumatoid arthritis risk loci identified by trans-ethnic GWAS meta-analysis in >100,000 subjects.

SNP	Chr.	Genes	A1/A2 (+)	Trans-ethnic		European		Asian	
				OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
rs227163	1	TNFRSF9	C/T	1.04 (1.02–1.06)	3.9E-04	1.00 (0.97–1.03)	9.3E-01	1.11 (1.08–1.16)	3.1E-09
rs28411352	1	MTF1-INPP5B	T/C	1.11 (1.08–1.14)	2.8E-12	1.10 (1.07–1.14)	5.9E-09	1.12 (1.06–1.19)	7.8E-05
rs2105325	1	LOC100506023	C/A	1.12 (1.08–1.15)	6.9E-13	1.12 (1.08–1.15)	3.3E-11	1.13 (1.04–1.23)	5.2E-03
rs10175798	2	LBH	A/G	1.08 (1.06–1.11)	1.1E-09	1.09 (1.06–1.12)	4.2E-08	1.07 (1.02–1.13)	6.4E-03
rs6732565	2	ACOXL	A/G	1.07 (1.05–1.10)	2.7E-08	1.10 (1.07–1.14)	9.4E-09	1.04 (1.00–1.08)	4.0E-02
rs6715284	2	CFLAR-CASP8	G/C	1.15 (1.10–1.20)	1.8E-09	1.15 (1.10–1.20)	2.5E-09	-	-
rs4452313	3	PLCL2	T/A	1.09 (1.06–1.12)	1.6E-10	1.11 (1.08–1.15)	5.2E-11	1.04 (0.99–1.09)	9.2E-02
rs3806624	3	EOMES	G/A	1.08 (1.05–1.11)	8.6E-09	1.08 (1.05–1.12)	2.8E-08	1.06 (0.99–1.14)	1.0E-01
rs9826828	3	IL20RB	A/G	1.44 (1.28–1.61)	8.6E-10	1.44 (1.28–1.61)	8.7E-10	-	-
rs13142500	4	CLNK	C/T	1.10 (1.07–1.13)	3.0E-09	1.10 (1.06–1.15)	2.4E-06	1.10 (1.04–1.15)	2.8E-04
rs2664035	4	TEC	A/G	1.07 (1.04–1.10)	9.5E-08	1.08 (1.05–1.11)	3.3E-08	1.03 (0.97–1.08)	3.3E-01
rs9378815	6	IRF4	C/G	1.09 (1.06–1.12)	1.7E-10	1.09 (1.05–1.12)	1.4E-07	1.10 (1.04–1.15)	2.3E-04
rs2234067	6	ETV7	C/A	1.15 (1.10–1.20)	1.6E-09	1.14 (1.09–1.19)	4.1E-08	1.22 (1.06–1.41)	7.0E-03
rs9373594	6	PPIL4	T/C	1.09 (1.06–1.12)	3.0E-09	1.07 (1.02–1.12)	6.5E-03	1.11 (1.07–1.15)	4.8E-08
rs67250450	7	JAZF1	T/C	1.10 (1.07–1.14)	3.7E-09	1.11 (1.07–1.14)	2.6E-09	1.02 (0.84–1.23)	8.5E-01
rs4272	7	CDK6	G/A	1.10 (1.06–1.13)	5.0E-09	1.10 (1.07–1.14)	1.2E-08	1.06 (0.98–1.15)	1.3E-01
rs998731	8	TPD52	T/C	1.08 (1.05–1.11)	1.9E-08	1.09 (1.06–1.12)	6.6E-09	1.02 (0.96–1.10)	4.9E-01
rs678347	8	GRHL2	G/A	1.08 (1.05–1.11)	1.6E-08	1.10 (1.06–1.13)	7.3E-09	1.03 (0.98–1.10)	2.6E-01
rs1516971	8	PVT1	T/C	1.15 (1.10–1.20)	1.3E-10	1.16 (1.11–1.21)	3.2E-11	-	-
rs12413578	10	10p14	C/T	1.20 (1.13–1.29)	4.8E-08	1.20 (1.12–1.29)	7.5E-08	-	-
rs793108	10	ZNF438	T/C	1.08 (1.05–1.10)	1.3E-09	1.07 (1.04–1.10)	6.1E-07	1.09 (1.04–1.14)	4.4E-04
rs2671692	10	WDFY4	A/G	1.07 (1.05–1.10)	2.8E-09	1.06 (1.03–1.09)	2.6E-05	1.10 (1.05–1.14)	9.9E-06
rs726288	10	SFTPD	T/C	1.14 (1.07–1.20)	1.6E-05	0.96 (0.86–1.06)	4.1E-01	1.22 (1.14–1.31)	8.8E-09
rs968567	11	FADS1-FADS2-FADS3	C/T	1.12 (1.07–1.16)	1.8E-08	1.12 (1.07–1.16)	1.8E-08	-	-
rs4409785	11	CEP57	C/T	1.12 (1.09–1.16)	1.2E-11	1.12 (1.08–1.16)	3.6E-09	1.16 (1.07–1.27)	4.3E-04
chr11:107967350	11	ATM	A/G	1.21 (1.13–1.29)	1.4E-08	1.21 (1.13–1.29)	1.1E-08	-	-
rs73013527	11	ETS1	C/T	1.09 (1.06–1.12)	1.2E-10	1.08 (1.05–1.11)	1.0E-06	1.14 (1.08–1.21)	4.1E-06
rs773125	12	CDK2	A/G	1.09 (1.06–1.12)	1.1E-10	1.09 (1.06–1.12)	2.1E-08	1.10 (1.04–1.17)	1.1E-03

SNP	Chr.	Genes	A1/A2 (+)	Trans-ethnic		European		Asian	
				OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
rs10774624	12	SH2B3-PTPN11	G/A	**1.09 (1.06–1.13)**	**6.8E–09**	**1.09 (1.06–1.13)**	**6.9E–09**	-	-
rs9603616	13	COG6	C/T	**1.10 (1.07–1.13)**	**1.6E–12**	**1.11 (1.07–1.14)**	**2.8E–11**	1.08 (1.02–1.14)	1.0E–02
rs3783782	14	PRKCH	A/G	**1.14 (1.09–1.18)**	**2.2E–09**	1.12 (0.96—1.31)	1.4E–01	**1.14 (1.09–1.19)**	**4.4E–09**
rs1950897	14	RAD51B	T/C	**1.10 (1.07–1.13)**	**8.2E–11**	**1.09 (1.06–1.12)**	**5.0E–08**	1.16 (1.08–1.25)	1.1E–04
rs4780401	16	TXNDC11	T/G	**1.07 (1.05–1.10)**	**4.1E–08**	**1.09 (1.06–1.13)**	**8.7E–09**	1.03 (0.98–1.08)	2.5E–01
rs72634030	17	C1QBP	A/C	**1.12 (1.08–1.17)**	**1.5E–09**	1.12 (1.06–1.19)	2.9E–05	1.12 (1.07–1.18)	9.6E–06
rs1877030	17	MED1	C/T	**1.09 (1.06–1.12)**	**1.9E–08**	1.09 (1.05–1.13)	1.3E–05	1.09 (1.04–1.14)	3.2E–04
rs2469434	18	CD226	C/T	**1.07 (1.05–1.10)**	**8.9E–10**	1.05 (1.02–1.08)	6.7E–04	**1.11 (1.07–1.15)**	**1.2E–08**
chr19:10771941	19	ILF3	C/T	**1.47 (1.30–1.67)**	**8.6E–10**	**1.47 (1.30–1.67)**	**8.8E–10**	-	-
rs73194058	21	IFNGR2	C/A	1.08 (1.05–1.12)	1.2E–06	**1.13 (1.08–1.18)**	**2.6E–08**	1.03 (0.98–1.08)	2.9E–01
rs1893592	21	UBASH3A	A/C	**1.11 (1.08–1.14)**	**7.2E–12**	**1.11 (1.07–1.15)**	**9.8E–09**	1.11 (1.05–1.18)	1.3E–04
rs11089637	22	UBE2L3-YDJC	C/T	**1.08 (1.05–1.11)**	**2.1E–09**	1.10 (1.06–1.15)	2.0E–07	1.06 (1.02–1.10)	8.9E–04
rs909685	22	SYNGR1	A/T	**1.13 (1.10–1.16)**	**1.4E–16**	**1.11 (1.08–1.15)**	**6.4E–12**	1.23 (1.14–1.33)	2.0E–07
chrX:78464616	X	P2RY10	A/C	**1.11 (1.07–1.15)**	**3.5E–08**	1.16 (0.78–1.75)	4.6E–01	**1.11 (1.07–1.15)**	**3.6E–08**

SNPs newly associated with $P < 5.0 \times 10^{-8}$ in the combined study of the Stage I GWAS meta-analysis and the Stage II and III replication studies of trans-ethnic (Europeans and Asians), European, or Asian ancestry are indicated. Association results with $P < 5.0 \times 10^{-8}$ are highlighted in bold. SNP IDs, positions, and alleles are based on positive strand of NCBI build 37. A1 represent risk allele of rheumatoid arthritis. Full results of the studies are indicated in Supplementary Table 1.

Chr., chromosome; freq., frequency; EUR, European; ASN, Asian; OR, odds ratio; 95%CI, 95% confidence interval; GWAS, genome-wide association study.