Identifying therapeutic targets for rheumatoid arthritis by genomics-driven integrative approaches

Jie Zhang1,4,5,6†; Xinyu Fang2,3†; Jingwei Wu2,3; Zixing Zhang2,3; Min Mu1,4,5,6*; Xinyu Fang2,3*, Dongqing Ye1,3,4,5,6*.

1Joint Research Center of Occupational Medicine and Health, Institute of Grand Health, Hefei Comprehensive National Science Center, Anhui University of Science and Technology, Hefei, Anhui, 231131, China
2Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, Hefei, Anhui, 230032, China
3Inflammation and Immune Mediated Diseases Laboratory of Anhui Province, Hefei, Anhui, 230032, China
4School of Public Health, Anhui University of Science and Technology, Hefei, Anhui, 231131, China
5Key Laboratory of Industrial Dust Prevention and Control ,Occupational Safety and Health, Ministry of Education, Anhui University of Science and Technology, Hefei, Anhui, 231131, China
6Anhui Institute of Occupational Safety and Health, Anhui University of Science and Technology, Hefei, Anhui, 231131, China

† Jie Zhang and Xinyu Fang contributed equally.
*Corresponding authors.

Correspondence to Professor Dongqing Ye, School of Public Health, Anhui

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
University of Science and Technology, Hefei, China; Email: ydqph@aust.edu.cn, Associated Professor Min Mu, School of Public Health, Anhui University of Science and Technology, Hefei, China; Email: candymu@126.com, and Associated Professor Xinyu Fang, Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, Hefei, China; Email: xinyufang@ahmu.edu.cn.

**Funding** This study was supported by Research Funds of Joint Research Center for Occupational Medicine and Health of IHM(OMH-2023-01;OMH-2023-08) and Anhui Province clinical medical research transformation project(NO.202304295107020041; NO.202304295107020048).
Summary

Genomics-driven drug discovery holds significant promise in identifying and developing novel therapeutic targets. Here, we utilized large-scale genomic data including genome-wide association studies (GWAS), rare variant burden tests in exome sequencing studies (Exome), and protein quantitative trait loci (pQTL), to prioritize therapeutic targets or repurpose drugs in rheumatoid arthritis (RA). We found that prioritized genes covering two approved RA treatment targets (IL6R and CD86), along with several targets currently undergoing active clinical trials for RA. Fifteen proteins were identified as having causalities with RA risk, three out of them showed strong support for colocalization. bromodomain-containing protein 2 (BRD2) was nominated as one of the most promising candidates for clinical translation as its wide expression in joint synovial tissues and validation in observational analyses associating with RA incidence. Collectively, our systematic prioritization of drug targets from different genetically informed approaches, and provided a comprehensive insight into therapeutic strategies for RA.

Keywords

GWAS, Exome, pQTL, gene prioritization, drug target discovery, Mendelian randomization, colocalization, rheumatoid arthritis
Introduction

Rheumatoid arthritis (RA), a common systemic autoimmune disease primarily characterized by chronic tissue inflammation and joint destruction,\(^1,2\) is highly heritable (heritability of RA has been estimated to be about 60 \%).\(^3\) In 2020, an estimated 17.6 million people had RA worldwide.\(^4\) Over recent decades, advancements in the treatment of RA have been notable, with the advent of biologic disease-modifying anti-rheumatic drugs (bDMARDs) and, more recently, targeted synthetic DMARDs.\(^5\) Nevertheless, despite the availability of these therapeutic drugs for RA, a considerable portion of RA patients fail to achieve low disease activity or remission.\(^6\) Additionally, there exists a substantial cost burden and disparities in these medications accessibility,\(^7,8\) further complicating the situation. Currently, screening for novel therapeutic targets remains a pressing need in the management of RA.

Using human genetic data to optimize the selection of targets has potential to lead to approved drugs.\(^9,10,11\) A previous genome-wide association study (GWAS) meta-analysis for RA with >100,000 subjects,\(^3\) served as a paradigm, underscoring the pivotal role of genetics in drug discovery efforts, which revealed drug target genes of approved RA drugs significantly overlapped with the 98 identified biological candidate genes, and indicated that CDK6/CDK4 inhibitors\(^12\) (approved drugs for different types of cancer) may hold promise for drug repositioning in RA treatment.

One constraint inherent in the singular approach of drug discovery driven by GWAS lies in the fact that a significant portion of GWAS signals stemming from common variants manifest negligible individual effect magnitudes. However, their potential
influence on molecular phenotypes, encompassing gene expression and protein abundance, might be significant to justify consideration for therapeutic intervention.\textsuperscript{13} Coupling large-scale GWAS with molecular quantitative trait loci (mQTL) datasets by utilizing Mendelian randomization studies, transcriptome-wide association studies, and colocalization methods provides insight to pathogenesis and facilitates future drug target prioritization.\textsuperscript{14,15,16,18,19} While GWAS frequently fail to elucidate gene function, analysis of whole-exome sequencing (WES) reveals rare coding variants, shedding light on novel biological functions for numerous genes and presenting potential avenues for therapeutic development.\textsuperscript{19,20} Due to the evolution of these genetically informed approaches, there has been a growing number of studies systematically assessing strategies to prioritize drug target genes.\textsuperscript{21,22} A recent study examined the efficacy of nominating drug targets from aforementioned genetically informed approaches (GWAS, QTL-GWAS, and Exome) across 30 clinical traits, and further integrated gene network diffusion to prioritize important drug target genes.\textsuperscript{21} Another cross-population meta-analysis study proposed a genomics-driven drug discovery framework comprising three complementary methodologies for gene prioritization, which successfully identified drug targets related to the coagulation cascade for venous thromboembolism.\textsuperscript{22} Here, we applied the core component of developed drug discovery framework\textsuperscript{22} to facilitate identification of therapeutic targets and repurposing opportunities for RA. We performed gene prioritization, overlap enrichment analysis, proteome-wide Mendelian randomization (MR), colocalization, candidate target validation analysis,
single-cell RNA-sequencing (scRNA-seq) analysis and population RA incidence association analysis by integrating the summary statistics from the large-scale RA GWAS, whole-exome sequencing and aptamer-based plasma protein measurements. Our findings provide insight into drug discovery for RA through the systematic integration of multi-layered genetics data.

**Results**

**Study overview and summary**

The study was designed and conducted as depicted in Figure1.

First, summary statistics were obtained from the largest large-scale GWAS for RA currently available,\textsuperscript{23} which were derived from 22,350 cases of RA and 74,823 controls in 25 European cohorts (in this study, all analyses were conducted exclusively on individuals of European ancestry, ensuring population homogeneity in genetic and other omics data). We utilized five gene prioritization tools by calculating gene scores or p values to seek for drug candidate genes,\textsuperscript{15,24,25,26,27} after that 84 candidate genes were jointly identified as high priority by four or more tools. Besides, we extracted gene-RA associations based on putative loss of function (pLOF) and deleterious missense variants with MAF <1\% from gene burden test results computed on WES data from the UK Biobank,\textsuperscript{19} which top 5\% significant genes (940 genes) were likewise treat as candidate genes. To identify opportunities for drug repurposing, we then assessed the overlap of candidate genes with targets of existing drugs and quantified an enrichment of the candidate genes in the target of clinical indication.
categories (i.e., Anatomical Therapeutic Chemical Classification System [ATC]).\textsuperscript{22,28} Out of these candidate genes, 21 were targets of pharmacological agents, which included two approved RA drug targets: $IL6R$ targeted by sarilumab and $CD86$ targeted by abatacept.

Secondly, we extracted plasma proteome GWAS data derived from the deCODE study, where plasma protein abundances were quantified utilizing 4,907 aptamers across a cohort comprising 35,559 individuals.\textsuperscript{17} Following this, we proceeded with proteome-wide MR analysis\textsuperscript{29} to estimate causal relationships between plasma proteins and RA. Through this analysis, we identified 18 proteins, including tumor necrosis factor (TNF-\(\alpha\)) and bromodomain-containing protein 2 (BRD2).

Subsequently, we validated these causal relationships using colocalization analysis.\textsuperscript{30} Among the proteins analyzed, hyaluronan and proteoglycan link protein 4 (HAPLN4), histone H1x (H1FX), and WNT1-inducible signaling pathway protein 1 (WISP1) exhibited high support evidence for colocalization.

Thirdly, we examined whether candidate target genes/proteins interacted with approved RA drug targets in the protein-protein interaction network\textsuperscript{31} and whether the signaling pathways they were significantly enriched in were related to the pathogenesis of RA\textsuperscript{32,33}. These validation analyses for utility of candidate target genes/proteins revealed that multiple prioritized genes (for example: colony stimulating factor 1 receptor, $CSF1R$) interactions with 7 targets of current RA medications, and significant enriched signaling pathways of these candidate targets were related to RA development (for example: JAK-STAT signaling pathways).
Finally, to elucidate the biological function of candidate targets for RA, we analyzed scRNA-seq data of joint synovial tissues samples from RA patients. The scRNA-seq analysis demonstrated that BRD2 widely expressed in joint synovial cells. Then verified whether the most promising target (BRD2) was associated with the onset of RA based on the UKB population cohort.

Each of these parts is described in more detail below.

**Gene prioritization and overlap enrichment in RA relevant medication categories**

we employed five gene prioritization tools in parallel, i.e., MAGMA, DEPICT, Priority index (Pi), Polygenic Priority Score (PoPS) and Transcriptome-wide association studies (TWAS). Briefly, MAGMA relies on a multiple linear principal components regression framework, integrating linkage disequilibrium (LD) among markers to identify multi-marker effects and ascertain gene significance through p-values.; DEPICT gene prioritization employs a phenotype- and mechanism-agnostic algorithm based on the assumption that truly associated genes share functional annotations, prioritizing genes through scoring, bias adjustment, and false discovery rate estimation steps; Pi represents a scoring system tailored for the drug development landscape within immune-related diseases, and it amalgamates various annotations, encompassing eQTL, chromatin interaction data, and genes associated with immune functions; PoPS leverages the entirety of the polygenic signal while incorporating comprehensive data regarding genes from diverse sources, such as cell-type-specific
gene expression and biological pathways, to conduct gene prioritization; TWAS calculates credible sets of causal genes through the utilization of prediction eQTL weights, LD, and GWAS summary statistics. The previous drug discovery framework has demonstrated that the comprehensive integration of five gene prioritization methods can effectively nominate candidate gene sets, thereby enhancing the enrichment of disease-relevant drug-target genes.²²

We observed significant enrichment of drug target genes (ORs: 2.74-3.88) in anti-tumor and immune modulator categories (L in ATC coding) using most of the five methods, with the exception of TWAS (Figure 2A, Table S1). This anti-tumor and immune modulator categories enrichment aligns well with the current mainstream drug categories for treating RA.³⁶ The omnibus results embracing 84 prioritized genes reach relatively higher enrichment (OR=5.19, p=0.002) (Table S1, Table S2). Notably, this omnibus integration nominated several drug targets including IL6R (targeted by tocilizumab/sarilumab, Interleukin-6 receptor alpha subunit inhibitor), which was previously proved to be a successful example of GWAS signals mapped to the target of an indicated pharmacological agent for RA (Figure 2B).³⁷ Likewise, there remained an edge-significant enrichment of drug target categories in anti-tumor and immune modulators (OR=1.58, p=0.07) due to that all prioritized genes from burden test not reach sufficient significance (all genes-RA associations with p > 1×10⁻⁵) (Table S3 & S4). Nevertheless, despite the limitations in statistical significance, this efficiently identified the target gene CD86, which corresponds to the clinically approved RA drug abatacept (Table S4).³⁶
Proteome-wide Mendelian randomization

Proteome-wide Mendelian randomization uses genetic predictors of protein level to gain insights into the putative causal implications of a multitude of proteins in influencing the risk of various diseases.\textsuperscript{38,39} For our proteome-wide Mendelian randomization analyses, we used cis-pQTLs located within a 1 Mb proximity to the transcription start site of the protein-coding gene as genetic instrument, because cis-pQTLs are considered to have a higher prior probability of specific biological effects.\textsuperscript{40} Following multiple test corrections, we identified 53 proteins that exhibited significant causality with RA (FDR < 0.05, Table S5). However, upon further examination using the heterogeneity in dependent instruments (HEIDI) test,\textsuperscript{41} distinguish proteins that were related with RA risk owing to a shared variant rather than genetic linkage, it was revealed that the causal association of 18 out of these 53 proteins with RA was not driven by linkage disequilibrium (\(P_{\text{HEIDI}} > 0.05\), Figure 3A & Table S5). The most prominent causal effect was observed with bromodomain-containing protein 2 (BRD2), where the odds ratio for RA was 26.8 (\(P_{\text{FDR}}= 3.03 \times 10^{-6}\)) per SD increase in genetically predicted levels (Table S5). Among the 15 MR-identified proteins associated with RA, 3 proteins (HAPLN4, H1FX and WISP1) had high support of colocalization analysis (\(PP_{H4} \geq 0.8\)) (Figure 3B,C & Table S5). \textit{HAPLN4} and \textit{WISP1} were also known as risk loci for RA.\textsuperscript{23} The co-expression of \textit{H1FX} with \textit{HMGB1}, as illustrated in Figure S1, the latter impacts overexpression of signal transduction receptors and abnormal regulation of
osteoclastogenesis and bone remodelling possibly linked with progression of RA. These 15 identified RA-associated proteins had networks, in particular physical interactions and co-expression (Figure 3D).

To verify the prioritized proteins within the framework of drug discovery, we proceeded to collate pharmaceutical compounds from the foremost drug repositories: DrugBank, Therapeutic Target Database, PharmGKB, and the Open Targets Platform. This yielded 22 drugs relevant to 3 protein-disease associations. (Table S6). In addition to TNF-α, a major inflammatory cytokine implicated in the pathogenesis of RA and targeted by several approved monoclonal antibody treatments (such as Infliximab, Etanercept, and Adalimumab), We also found that literature-based supportive evidence for biologic effects of BRD2 and GPNMB on RA were aligned with putative causality by MR estimates: For instance, Inhibition of BET family proteins (for example: BRD2) suppresses the inflammatory, matrix-degrading, proliferative and chemoattractive properties of rheumatoid arthritis synovial fibroblasts. Glycoprotein nonmetastatic melanoma protein B (GPNMB) expression in the majority of synovial fibroblasts and tissue-infiltrating monocyte for rheumatoid arthritis and osteoarthritis.

**Integrative validation for drug repurposing opportunities**

Integrative approaches have substantial efficiency in screening candidate drug targets: the 21 drug-target genes and 3 actionable proteins in total was nominated (Figure 4A). Among the eight drug targets identified, three (TNF, IL6R, and CD86) were targeted
by approved RA medications, while five (CD28, CD19, CSF1R, ESR1, and IFNAR1) were the focus of drugs undergoing clinical trials for RA (Table 1), thus bolstering the validity of genomics-guided drug discovery and repositioning for RA.

The PPI network revealed enrich interactions between candidate targets and the seven targets of current RA medications (TNF, CD86 and IL6R were also prioritized targets in our study) (Figure 4B & Figure S2). JAKs-MTOR, MET, CDK6, IL2R, FGF2 and IFNAR1 were determined to have known interactions. For example, JAK1 can be phosphorylated by by four cytokine-receptor families such as IFNAR1 and IL-2R, which have been validated as pathogenic pathways in RA.

We conducted further evaluation of the pathway enrichment of the candidate targets and observed significant enrichment for 150 pathways from Reactome (Table S7), as well as 20 from KEGG (Figure S3). Relevant signaling pathways implicated in the pathogenesis of RA were validated, including the JAK-STAT and PI3K/AKT signaling pathways. Particularly noteworthy was the transcriptional regulation by RUNX3 signaling pathway (FDR= 3.26 × 10^-5), which was nominated by integrating three drug-discovery approaches (Figure 5). This pathway included five genes separately prioritized by the three approaches: PSMB8 and PSMB9 by the omnibus gene prioritization approach, BRD2 by proteome-wide MR, and PSMB2 and PSMB5 by the gene burden tests from WES. RUNX3 signaling pathway might play roles on the mechanisms of T cell activation in RA. Our finding demonstrates that the three approaches synergistically prioritized therapeutic targets for RA. We also observed BRD2, PSMB2, PSMB5, PSMB8 and PSMB9 were almost moderate to high
expression in most human tissues (Figure S4).

**Follow-up analyses for the putatively therapeutic target**

To explore whether the candidate targets had any cell type-specific enrichment in joint synovial tissues, we further performed single cell-type expression analysis using data from 5,265 scRNA-seq profiles of 36 RA patients (Single-Cell Portal of the Broad Institute, AMP Phase 1 study). Our observations indicated that BRD2 exhibits a notably widespread expression across various cell types in joint synovial tissues, especially when compared to other candidate targets (Figure 6A). The expression values of BRD2 in each cell, positioned in a t-Distributed Stochastic Neighbor Embedding (tSNE) plot, and relatively higher expression values of BRD2 were found in B cells and T cells (Figure 6B). These cells correspond to 18 clusters within 5 cell types, namely fibroblasts, B cells, monocytes, plasmablasts, and T cells (Figure 6C). Specifically, we found that BRD2 was significantly expressed in 2 types of B cells ( naïve IGHD+CD27- and autoimmune-associated B cells), plasmablast, 3 types of fibroblasts (HLA-DRAhi sublining, DKK3+ sublining and CD55+ lining fibroblasts), NUPR1+ monocytes, 4 types of T cells (CCR7+ and three CD8+ clusters: GZMK+, GNLY+GZMB+ cytotoxic lymphocytes, and GZMK+GZMB+ T cells) (permutation test $P < 0.001$; Figure 6D). We also found BRD2 differential expression between rheumatoid arthritis patients ($n=315$) and healthy control ($n=315$) in CD14dim.CD16+ cell from Rheumatoid Arthritis Bioinformatics Center$^{52}$ (RABC, dataset id: RABC18) (Figure S5).
Given that BRD2 was identified as one of the highest confidence targets supported by diverse approaches, we sought to investigate whether BRD2 might have potentially beneficial or deleterious effects on traits beyond RA. To accomplish this, we screened the results from a phenome-wide association study (PheWAS) performed on the AstraZeneca PheWAS Portal.53 While BRD2 itself was not significantly associated with any other traits at the gene level (Figure S6), an indel variant (6-32977921-TGAG-T) of BRD2 was found to be positively associated with psoriasis and psoriatic arthropathy, which the odds ratios (ORs) and p-values for these associations were 1.95 (9.12 × 10^{-122}) and 2.50 (6.36 × 10^{-36}), respectively. Furthermore, a missense variant (6-32974662-A-G) of BRD2 was observed to be associated with multiple sclerosis, with an OR of 9.30 (P = 0.0052) (Figure 6E & Table S8). These evidence suggesting that inhibition of bromodomain proteins have potentiality in drug reposition for rheumatic diseases.54

To triangulate the evidence regarding the effect of BRD2 on RA, we further evaluated whether rs206783 (the top cis-pQTL of BRD2) was associated with RA incidence in the general population, such as in the UK Biobank dataset. Our analysis revealed that individuals with the genotype (AA) of rs206783 exhibited a slightly higher risk of RA, with an odds ratio (OR) of 1.10 (95% CI: 1.01-1.20), and a p-value of 0.03, when compared to those with the genotype (CC) (Figure 6F). The significant relevance between rs206783 and RA was reaffirmed in the FinnGen study using RA GWAS summary statistics rather than population observational data (Table S9).
Discussion

In the present study, we integrated three genomics-driven approaches (GWAS, QTL-GWAS, and Exome) to prioritize drug target genes/proteins. We further evaluated their biological and clinical relevance with RA to ascertain their utility in discovering candidate drug targets. GWAS and exome analysis complement each other in prioritizing genes for drug targeting when considering overlap with drug target in RA-relevant medication codes (primarily anti-tumor and immune modulator categories). Of note, both approaches corroborated IL6R (targeted by approved RA drug sarilumab) and CD28 as candidate drug targets for RA. Whereas currently, only a relatively few clinical trials have proved the abilities of remaining candidate targets such as IFNAR1 antagonist [Anifrolumab] and CSF1R Inhibitor [JNJ-40346527] in achieving remission of RA. Furthermore, proteome-wide MR improved drug development yield by revealing causality of 15 proteins with RA beyond gene prioritization. The two causal proteins (BRD2 and GPNMB) have the potential as therapeutic targets for RA as indicated in the supporting literatures.\textsuperscript{47,48} The subsequent single-cell RNA sequencing analysis revealed high expression of BRD2 in immune cells and fibroblasts, indicating its possible role in inflammation and fibrogenesis, thereby triggering the development of RA. Finally, we assessed the distinct effects of BRD2 on wide-range traits through phenome-wide association studies (PheWAS) and found its top \textit{cis}-pQTL (rs206783) significant association with RA incidence at the population level.

The gene priority determined by the Exome method mostly differs from that
determined by the GWAS method, a distinction also reflected in the identified drug targets. The odds ratio of drug target enrichment in the anti-tumor and immune modulator medicine categories for the Exome method was 1.58 ($P=0.068$), which was significantly lower than that of the GWAS method (OR 5.19, $P=0.002$). This difference may be attributed to two situations: Firstly, the Exome method exclusively focuses on rare variations with a minor allele frequency (MAF) less than 1% in the WES data. The gene score is derived from the gene load test of functional loss and missense variation, which may lead to relatively low power, a trend consistent with previous studies comparing gene prioritization methods. Secondly, gene prioritization based on the GWAS method is an omnibus strategy (Contains five gene prioritization methods with respective advantages), and one of its components, PI, could lead to a considerably higher enrichment in immune drugs, as consistently reflected in our results.

The confirmation of CD28 as a candidate drug target by both GWAS and exome analysis is notable. CD28 superagonists (CD28SA), exemplified by TGN1412, is a CD28-specific monoclonal antibody known to activate T cells without substantial TCR involvement. However, early-phase I clinical trials of TGN1412 encountered severe cytokine storms, prompting termination of development. Subsequent investigations suggested that appropriate dosages, such as those used for TAB08 (another CD28 superagonists), could selectively activate regulatory T cells without inducing significant release of pro-inflammatory cells (Clinical trials identifie: NCT01885624). Phase Ib trials in RA patients demonstrated acceptable adverse
events level associated with TAB08 (Clinical trials identifier: NCT01990157). Future phase II trials are needed to further establish TAB08's capacity to activate regulatory T cells for immune diseases. Colony-stimulating factor 1 receptor (CSF1R) emerged as another promising drug candidate target in our study. CSF1R plays a crucial role in modulating monocyte proliferation, migration, and activation, all of which are considered significant factors in the pathogenesis of RA. Increased expression of CSF1R has been observed in the synovium of RA patients, and blockade of CSF1R has demonstrated protective effects against bone and cartilage destruction in mouse models. However, the results of a phase IIa parallel-group study (Clinical trials identifier: NCT01597739) investigating the oral CSF-1R inhibitor JNJ-40346527 in patients with active rheumatoid arthritis did not show a significant improvement in disease activity or inflammatory levels. These studies collectively indicate that there remains a substantial gap between genetic-supported drug candidates and their actual clinical applications, underscoring the need for further in-depth research and verification.

Mendelian randomization contribute to genetic drug target validation, which utilized instrument variable to proxy drug target and estimated causal relation with diseases. Considering that proteins are typically the direct targets of drugs in most cases (over 90% of drug targets are proteins), protein exposure is usually the preferred choice for drug target MR. A prior investigation assessed the causal effects of inflammation-related proteins across immune-mediated disorders, elucidating the directionally disparate impact of CD40 on the susceptibility to RA compared to other
autoimmune conditions such as multiple sclerosis and inflammatory bowel disease.\textsuperscript{15} While we were unable to replicate the causal association between CD40 and RA in proteome-wide MR analysis due to the unavailability of CD40 pQTLs for present MR testing, our study did demonstrate that four out of the five gene prioritization methods based on GWAS jointly confirmed its prioritization. Furthermore, in situations where genetic instrument variables for related proteins are unavailable, upstream markers such as mRNA or downstream markers can also serve as indirect confirmations of alternative exposure.\textsuperscript{58} A recent study investigated the association between cis-expression quantitative trait loci (cis-eQTL) and RA risk using MR and colocalization analysis as the primary study design. The study identified CCR6, HLA-DPA1, HLA-DRB1, IFNGR2, C5, ATP2A1, and FEN1 as potential candidate drug target genes, and candidate drug prediction and molecular docking were utilized for further verification.\textsuperscript{59} We study used proteome-wide MR analysis based on cis-pQTL and obtained a list of candidate drug targets that were completely different from this study, this discrepancy indicated the MR estimation of tissue-specific drug targets through eQTL weighted analysis may not always align with evidence from similar blood-based pQTL analysis or drug trials.\textsuperscript{58} While our study and existing research did not yield shared determined actionable subjects through drug target MR, these findings provide valuable insights into drug target prioritization and medicinal potential from different perspectives. Among the 15 causal proteins identified by our study, three of them (HAPLN4, H1FX, and WISP1) were verified to have colocalization support with RA. However, little is known about the potential role of
these three proteins in the pathogenesis of RA. H1FX may contribute to the induction and persistence of complement activation and arthritis in RA, as it is one of many citrullinated peptides found in the synovial fluid of RA patients, which are targets of the anti-citrullinated protein antibody response. Genetic polymorphisms in WISPI have been associated with RA susceptibility. WISPI plays a significant role in regulating Wnt signaling in bone homeostasis, which might be linked to the progression of RA. HAPLN4 has been identified as a drug-targeting gene for schizophrenia and bipolar disorder. Additionally, existent evidence have shown consistent effect directions of multiple genetic variants between schizophrenia-RA and bipolar disorder-RA pairs. These findings suggest that HAPLN4 may serve as a therapeutic target not only in psychiatric disorders but also in immune disorders like RA. This indicates potential shared mechanisms or pathways underlying these seemingly distinct indications, opening avenues for further exploration into their treatment.

The series of results from our current study suggest promising clinical prospects for BRD2 in the treatment of RA. BRD2 is belong to the bromodomain and extra-terminal (BET) family member proteins. BRD2, BRD3, and the most studied BET protein BRD4 are ubiquitously expressed in mammalian cells and tissue types. A recent investigation has illustrated that targeted suppression of the second bromodomains within BET proteins (referred to as iBET-BD2) efficiently attenuated the synthesis of pivotal pro-inflammatory mediators, notably Th17 cytokines, within a co-culture system comprising B and T cells. Remarkably, this inhibition exerted no
discernible influence on the proliferative capacity of the cells, yet it notably diminished the secretion of effector cytokines, including IFN-γ, IL-17A, and IL-22. Subsequently, the study tested an optimized iBET-BD2 compound (GSK620), which showed a significant dose-dependent inhibition of both arthritic score and IgG1 production in response to immunization in a collagen-induced arthritis model in rats. Additionally, BET inhibitors influence the inflammatory responses of various immune cells, including B cells, macrophages, synovial fibroblasts, and chondrocytes, indicating a broad role of BRD proteins in autoimmune disorders. These findings align with the anticipated effects of inhibitors for BRD2 on RA. The extensive and high expression of BRD2 observed in B cells, T cells, and fibroblasts in joint synovial tissues, as identified in our current study, suggests a potential role for BRD2 inhibition in modulating the inflammatory response associated with RA.

In summary, our study has facilitated the in silico screening of abundant drugs and targets with supporting evidence. It underscores that thorough evaluation of genetic evidence offers indispensable potential in advancing clinical therapy for RA.

**Limitations of the study**

Several limitations should be considered. First, the primary focus of this study lies in examining the overlap and enrichment of prioritized genes with the drug targets in anti-tumor as well as immunosuppressive medications, potentially overlooking effective targets within alternative categories. Second, despite encompassing a wide array of blood proteins in this study, there is a possibility that we have inadvertently
neglected significant proteins lacking genetic instruments. Moreover, besides proteins, other intermediary molecular quantitative traits (eQTL, mQTL, sQTL) may offer further insights. Third, subtle difference in population structure between pQTL (Only Iceland population) and GWAS (Mix European population). Fourth, for the observational analysis and verification of the association between BRD2 and RA, we did not directly utilize the standard expression of BRD2 (as measured by the UKB through the olink platform). This decision stemmed from the relatively small number of new RA cases within the population with available standard protein expression data for BRD2, potentially resulting in insufficient statistical power. Finally, while we have substantiated some prospective targets by scrutinizing evidence derived from biological efficacy or clinical trials, it is imperative that direct in vivo or in vitro experimentation be undertaken as the subsequent phase for the prioritized targets identified in this investigation.

Acknowledgments
We acknowledge investigators and participants in the UK Biobank study, the deCODE project and cited genome-wide association studies for sharing data. The authors also thank Dr. Xianyong Yin for guide on statistical methodology. This study was supported by Research Funds of Joint Research Center for Occupational Medicine and Health of IHM(OMH-2023-01;OMH-2023-08) and Anhui Province clinical medical research transformation project (NO.202304295107020041; NO.202304295107020048)
**Author contributions**

Jie Zhang conceived and designed the study. Jie Zhang and Xinyu Fang performed statistical analyses. Xinyu Fang provided guidance on statistical analyses. Dongqing Ye supervised all statistical analyses. Jie Zhang and Xinyu Fang contributed to the collection and interpretation of data. All the authors contributed by providing advice on interpretation of results. Jie Zhang drafted the manuscript. All authors read, approved, and provided feedback on the final manuscript.

**Declaration of interests**

The authors declare no competing interests

**Figure Legend**

**Figure 1. Study overview and summary**

We sought therapeutic targets and repurposing opportunities for rheumatoid arthritis by conducting integrative genomics-driven approaches: we conducted gene prioritization based on GWAS and Exome summary statistics, evaluating whether the prioritized genes overlapped with and were enriched in drug targets within the RA-relevant medication category. Concurrently, we identified target proteins through proteome-wide Mendelian randomization and colocalization analysis using pQTL summary statistics. Secondly, we performed protein-protein interaction analysis between the selected candidate drug targets and existing approved RA drugs to explore
the potential for drug repurposing. Additionally, we conducted enrichment analysis to assess the biological rationale for their use in RA treatment. Finally, we conducted single-cell sequencing, Phewas, and observational analysis on one of the plausible drug targets, BRD2, to underscore the clinical translational prospects of developing RA therapies linked with it. (This figure created with BioRender.com)

**Figure 2. Gene prioritization and overlap enrichment in various medication categories**

(A) Overall enrichment of drug-target genes nominated by five gene prioritization tools and their omnibus results. The error bars represent 95% confidence intervals.

(B) Enrichments for the omnibus results per disease and ATC code. OR, odds ratio;

**Figure 3. Identification of casual proteins by proteome-wide Mendelian randomization and colocalization analysis**

(A) Volcano plot for associations of genetically predicted 1881 plasma proteins levels with RA in MR analysis. Labelled and colour genes refer to MR findings with FDR-corrected $P<0.05$ and $P_{\text{HEIDI}} > 0.05$. Red genes indicate the positive effect of the plasma proteins on RA; blue genes indicate the negative effect of the plasma proteins on RA.

(B) Evidence for colocalization between plasma proteins and RA. Circle size indicates the colocalization $P$ value for H4 and the colour of the circle indicate the classification of the evidence. Evidence support order defined as: $PP_{H4} < 0.5$ no colocalization support; $PP_{H4} 0.5-0.8$ medium colocalization support; $PP_{H4} >0.8$ high colocalization support.

(C) High support for colocalization with 3 plasma proteins including HAPLN4, H1FX and
WISP1 (Presented sequentially from left to right)

(D) Networks of identified causal proteins associated with RA. The network prediction was based on GeneMANIA (http://www.genemania.org).

**Figure 4. Comprehensive evaluation of drug reposition potentials of nominated candidate drug targets for RA treatment.**

(A) Candidate target genes nominated for rheumatoid arthritis drug discovery. The colors on the Manhattan chart denote various nomination methods, while the circle type signifies the nominated drug targets at different stages of RA treatment (including approved drugs, those in clinical trials, and targets yet to be clinically tested in RA treatment).

(B) Protein-Protein interaction between 6 approved RA medications (7 drug targets) and nominated candidate drug targets. The blue color of nominated candidate drug targets denotes that the type of interaction is known (determined from curated datasets or experiments).

**Figure 5. The three drug discovery approaches complementarily prioritized drug targets for RA**

The genes prioritized for RA by the three drug discovery approaches were connected to each other if their protein-protein interaction scores were larger than 0.4. Line thickness represents the strength of protein-protein interaction.

**Figure 6. Multiple layers of evidence support the clinical efficacy of putative drug targets in RA treatment.**
(A) Candidate drug target genes expression levels of each cell type at single-cell resolution in the joint synovial tissues of 36 RA patients.  

(B) BRD2 expression levels of each cell type  

(C) 18 annotated cell types clusters across 5,265 cells from all cell types in the joint synovial tissues.  

(D) BRD2 expression status in 5,265 scRNA-seq profiles, including 1,142 B cells, 1,844 fibroblasts, 750 monocytes, and 1,529 T cells. BRD2 expression levels of 18 cell types are shown in the box plot. In each box, the horizontal solid line represents the median of the expression level, and the horizontal dashed line represents the average value. Each box extends from the 25th percentile of each group to the 75th percentile. Whiskers extend 1.5 times the interquartile distance from the top and bottom of the box.  

(E) Manhattan plot for PheWAS of BRD2 at variant level. 35 variant-trait pairs reach the significant genome-wide association threshold ($P < 5 \times 10^{-8}$), 53 variant-trait pairs reach the suggestive threshold ($P < 5 \times 10^{-6}$). Upward triangle represents positive association, downward represents negative association.  

(F) Cumulative risk of RA incidence between different rs206783 genotype (AA, AC, CC) in UK Biobank population (424,282 individuals, 4897 incident RA cases). rs206783 is the index $cis$-pQTL of BRD2.
Table 1. Review of RA-related randomized controlled trials on candidate targets prioritized by integrative approaches

<table>
<thead>
<tr>
<th>Study</th>
<th>NCT number</th>
<th>Drug</th>
<th>Target</th>
<th>Pharmacological action</th>
<th>Condition</th>
<th>Clinical endpoint</th>
<th>Intervention</th>
<th>N</th>
<th>Estimation parameter and value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpublished</td>
<td>NCT01990157</td>
<td>TAB08</td>
<td>CD28</td>
<td>specific surface glycoprotein CD28 agonist/T-cell-specific surface glycoprotein CD28 agonist</td>
<td>active RA</td>
<td>Incidence of dose-limiting toxicity in every dose cohort</td>
<td>Multiple TAB08 administrations as intravenous infusions.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unpublished</td>
<td>NCT00639834</td>
<td>MDX-1342</td>
<td>CD19</td>
<td>anti-CD19 fully human monoclonal antibody</td>
<td>active RA</td>
<td>incidence and severity of treatment-emergent adverse events</td>
<td>One dose of active MDX-1342 will be administered to patients as an i.v. infusion.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Genovese MC, 2015</td>
<td>NCT01597739</td>
<td>JNJ-40346527</td>
<td>CSF1R</td>
<td>CSF-1R Inhibitor</td>
<td>active RA</td>
<td>Change from baseline in the Disease Activity Score (DAS28), using C-reactive protein (CRP) [Time Frame: Week 12]</td>
<td>100 mg capsule, OD, BID</td>
<td>95 patients (63 JNJ-40346527, 32 placebo)</td>
<td>Mean improvements in DAS28-CRP from baseline to Week 12 were 1.15 for the JNJ-40346527 group and 1.42 for the placebo group</td>
<td>0.3</td>
</tr>
<tr>
<td>Cho SK, 2021</td>
<td>NCT02602704</td>
<td>bazedoxifene</td>
<td>ESR1</td>
<td>Estrogen receptor modulator</td>
<td>RA</td>
<td>bone mineral density (BMD) change in the lumbar spine (L-spine) from baseline to 48 weeks</td>
<td>Bazedoxifene 20 mg/day (Viviant) &amp; Elemental calcium 1200mg daily and vitamin D 800 IU daily (Caltrate D 400 * 2/day)</td>
<td>114 patients (57 patients in each group)</td>
<td>increase in L-spine BMD (0.015 g/cm2 ) was observed in the bazedoxifene group</td>
<td>0.007</td>
</tr>
<tr>
<td>Unpublished</td>
<td>NCT03435601</td>
<td>Anifrolumab</td>
<td>IFNAR1</td>
<td>Interferon-alpha/beta receptor alpha chain antagonist</td>
<td>RA patients with an increased type I IFN gene signature</td>
<td>Achieving an ACR 20 response at week 24</td>
<td>IV Administration of Anifrolumab 300 mg every 4 weeks from week 0 to week 20 for a total of 6 doses.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
## STAR Methods

### Key resources table

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK Biobank</td>
<td>UK Biobank</td>
<td><a href="https://www.ukbiobank.ac.uk/">https://www.ukbiobank.ac.uk/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software and algorithms</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MAGMA</td>
<td>Leeuw et al., 2015</td>
<td><a href="https://ctg.cnrc.nl/software/magma">https://ctg.cnrc.nl/software/magma</a></td>
</tr>
<tr>
<td>DEPICT</td>
<td>Pers et al., 2015</td>
<td><a href="https://data.broadinstitute.org/mpg/depict/">https://data.broadinstitute.org/mpg/depict/</a></td>
</tr>
<tr>
<td>Pi</td>
<td>Fang et al., 2019</td>
<td><a href="http://bioconductor.org/packages/release/bioc/html/Pi.html">http://bioconductor.org/packages/release/bioc/html/Pi.html</a></td>
</tr>
<tr>
<td>PoPS</td>
<td>Weeks et al., 2020</td>
<td><a href="https://github.com/FinucaneLab/pops">https://github.com/FinucaneLab/pops</a></td>
</tr>
<tr>
<td>FOCUS</td>
<td>Mancuso et al., 2019</td>
<td><a href="https://github.com/bogdanlab/focus">https://github.com/bogdanlab/focus</a></td>
</tr>
<tr>
<td>GREP</td>
<td>Sakaue et al., 2019</td>
<td><a href="https://github.com/saorisakaue/GREP">https://github.com/saorisakaue/GREP</a></td>
</tr>
<tr>
<td>SMR</td>
<td>Zhu et al. 2016</td>
<td><a href="https://yanglab.westlake.edu.cn/software/smr/#SMR&amp;HEIDI">https://yanglab.westlake.edu.cn/software/smr/#SMR&amp;HEIDI</a> analysis</td>
</tr>
<tr>
<td>Coloc</td>
<td>Giambartolomei et al., 2014</td>
<td><a href="https://cran.r-project.org/web/packages/coloc/index.html">https://cran.r-project.org/web/packages/coloc/index.html</a></td>
</tr>
<tr>
<td>PheWAS</td>
<td>AstraZeneca PheWAS Portal</td>
<td><a href="https://azphewas.com/">https://azphewas.com/</a></td>
</tr>
</tbody>
</table>

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license. It is made available under a permanent [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dongqing Ye (ydqph@aust.edu.cn).

Materials availability

This study did not generate new unique reagents.

Method details

Genetic summary statistics resources

GWAS summary statistics for rheumatoid arthritis (RA) were obtained from the Ishigaki K, et al’s study with 22,350 cases of RA and 74,823 controls of European ancestry. The summary statistics data can be accessed via the GWAS Catalog using the access ID: GCST90132223.

We utilized gene burden test outcomes derived from whole-exome sequencing (WES) data obtained from the UK Biobank. Gene-RA associations were extracted, focusing on putative loss-of-function (pLOF) and deleterious missense variants with a minor allele frequency (MAF) below 1%. These associations were available for approximately 18,800 genes, ranked by association p-value, and retrieved using the provided Ensembl identifier.

Ferkingstad et al., 4719 proteins measured in 35,559 Icelanders, were employed to
extract summary statistics of genetic associations with plasma proteins.\textsuperscript{17} We extracted the \textit{cis}-pQTL variants under multiple-testing significance threshold (determined using Bonferroni correction for all 27.2 million tested variants, giving a threshold of $0.05/27,200,000=1.8 \times 10^{-9}$), and \textit{cis}-pQTL was defined as a SNP residing within 1 megabase (Mb) upstream or downstream of the transcription start site (TSS) of the corresponding protein-coding gene.

\textbf{Gene prioritization}

We used five distinct tools for gene prioritization based on RA GWAS summary statistics, namely MAGMA\textsuperscript{24}, DEPICT\textsuperscript{25}, Priority Index (Pi)\textsuperscript{15}, Polygenic Priority Score (PoPS)\textsuperscript{26}, and Transcriptome-wide Association Studies (TWAS)\textsuperscript{27}. We analyzed with default settings unless explicitly specified otherwise.

MAGMA is a gene analysis method that accounts for LD structure, which projects the SNP matrix for a gene onto its principal components (PC), obtained PCs without small eigenvalues as predictors to examine association between genes and phenotype in the linear regression model. For MAGMA,\textsuperscript{24} “gene-model snp-wise = mean” option was set. DEPICT operates by taking a collection of trait-associated SNPs as input, leveraging them to detect independently associated loci, which may encompass multiple genes. DEPICT predicts the biological functions of genes across a broad spectrum of biological annotations, comprising 14,461 gene sets. Subsequently, basing on this information of gene sets, DEPICT identifies reconstituted gene sets that exhibit enrichment for genes within the associated loci, and prioritizes genes within
associated loci by identifying genes from different loci that share similar predicted functions. We selected these variants with relaxed p-values $< 1.0 \times 10^{-5}$ as input for DEPICT.\textsuperscript{25} Pi is primarily constituted by a gene-predictor matrix that integrates genomic and annotation predictors, aiming to prioritize approximately 15,000 genes for a given trait. Initially, Pi identifies "Seed genes" by leveraging scores from genomic predictors to ascertain functional linkage to the input disease-associated genetic variant, considering factors such as proximity, conformation, and expression. Subsequently, scores from annotation predictors related to immune function, phenotype, or disease are exclusively applied to these seed genes. Non-seed genes are determined based on knowledge of network connectivity. The predictor matrix generates a numerical Pi prioritization rating ranging from 0 to 5, along with corresponding rankings, while affinity scores ensure comparability across different predictors for both seed and non-seed genes. Following the methodology of the original study,\textsuperscript{15} we utilized lead variants with a significance threshold of $p < 5.0 \times 10^{-8}$ as input, eQTL in the peripheral blood and immune cells, chromatin interaction in immune cells, topologically associating domain boundary in the GM12878 cell line, and the STRING\textsuperscript{31} protein-protein interaction network (PPI), with a high confidence score. PoPS is a tool that attributes high scores to genes exhibiting features akin to those with robust gene-level associations. It ran under the assumption that genes located proximally to associated SNPs and sharing analogous biological annotations are more likely to be causally related. PoPS utilizes gene-level associations derived
from GWAS summary statistics to ascertain collective polygenic enrichments of gene features sourced from cell-type-specific gene expression, biological pathways, and PPI. In order to nominate potential causal genes, PoPS assigns a priority score to each protein-coding gene based on these enrichments. In this context, we opted to employ top-ranked genes rather than pinpointed genes, thereby incorporating multiple genes per locus for drug discovery purposes, as the conduction done by previous study. We conducted FOCUS to disentangle the causal and tagging gene-trait associations at a TWAS-significant region, which FOCUS serves as a fine-mapping tool, estimates credible sets of causal genes by integrating prediction eQTL weights, LD, and GWAS summary statistics. We iteratively executed FOCUS across 44 GTEx v7 tissues, and calculated the median of their posterior inclusion probabilities (PIP).

We prioritized genes according to predefined thresholds as follows: for MAGMA, genes were prioritized if they exhibited a FDR < 0.05; for Pi and PoPS, the top 5% of genes were selected based on descending gene scores; for DEPICT, a more inclusive threshold of FDR < 0.2 was applied, as DEPICT calculates p-values exclusively for genes within pre-featured target loci by default; and finally, genes were prioritized from TWAS results if they demonstrated a PIP > 0.1.

**Proteome-wide Mendelian randomization (MR) analysis**

Summary-data-based Mendelian randomization (SMR) analysis was conducted to estimated the causal associations between proteins and RA. We used the following settings: --maf 0.01, --thread-num 10, --diff-freq 1. To control for genome-wide type I
error, we conducted FDR correction.

To distinguish proteins that were related with RA risk owing to a shared variant rather than genetic linkage, we further examine using the heterogeneity in dependent instruments (HEIDI) test. The p value of the HEIDI test > 0.05 indicated that the association of protein and RA was not driven by linkage disequilibrium.

**Colocalization analysis**

We employed a Bayesian colocalization approach to identify shared genetic variants between RA GWAS signals and cis-pQTL using “coloc” R package. We used the default coloc priors for Bayesian colocalization analysis, with \( p_1 = 1 \times 10^{-4} \) (prior probability a SNP is associated with protein), \( p_2 = 1 \times 10^{-4} \) (prior probability a SNP is associated with RA), and \( p_{12} = 1 \times 10^{-5} \) (prior probability a SNP is associated with both protein and RA). For each protein (establishing causality with RA determined by our SMR analysis), we extracted the region located within 1 Mb around the cis-pQTL.

Five posterior probabilities (PPs) were calculated for the colocalization analysis: \( PP_0 \) (null model of no causal genetic variant); \( PP_1 \) (probability that causal variants only associated with protein); \( PP_2 \) (probability that causal variants only associated with RA); \( PP_3 \) (probability that RA and proteins causal variants are distinct), and \( PP_4 \) (probability that RA and proteins share same causal variants). The genes were defined as high support for colocalization with cut-off of \( PP_4 \geq 0.8 \). The visualization of colocalization is achieved through ezQTL website based tool.

**Function and network prediction**
We utilized GeneMANIA to predict the biological network and functions of cis-genes associated with RA-related proteins. Further details regarding the datasets incorporated into GeneMANIA can be found elsewhere. To explore the potential interactions between identified candidate target and current RA medication targets, a PPI network was constructed using the STRING database (https://string-db.org/), we retained the protein-protein pairs with interaction scores larger than 0.4.

**Clinical evidence evaluation**

We undertook a systematic review focusing on RA-related clinical trials targeting candidate targets. Our search encompassed studies across three databases: MEDLINE, EMBASE, and the clinical trials registration database, with publications up to February 30th, 2023. Trials that were not randomized controlled trials (RCTs) or not conducted on human subjects were excluded from our analysis. Key information, including the first author, year of study, National Clinical Trial number, patient characteristics, sample size, intervention, trial phase, trial status, assessment of efficacy, and adverse effects, were systematically extracted from eligible studies.

**Single cell-type expression analysis**

To disentangle the BRD2 expression pattern in the joint synovial tissues of RA patients, we obtained 5,265 scRNA-seq profiles of 36 RA patients from Zhang, F. et al study from the Single-Cell Portal of the Broad Institute (https://singlecell.broadinstitute.org/single_cell/) (accession ID SCP279).
We reanalyzed the data, focusing on BRD2 expression status. We analyzed the gene expression matrix and the associated metadata using R v.4.1.2 and “Seurat” R package (https://satijalab.org/seurat/). To cluster the cells, we adopted the clustering annotation from the original study. To test whether BRD2 expression was enriched in specific cell types, we calculated the proportion of BRD2-expression cells in this cell type. Subsequently, we permuted the cell type labels 1,000 times and obtained the frequency (permutation P value) of the same cell type containing the same or a larger proportion of BRD2-expression cells. Additionally, we compared the BRD2 expression level between a target cell type and the other cell types using Wilcoxon rank-sum test.

**UK Biobank genotyping data and quality control**

The genotypes of the UK Biobank participants were assayed using either of two genotyping arrays, the Affymetrix UK BiLEVE Axiom (approximately 50,000 participants) or Affymetrix UKBiobank Axiom genotyping array (approximately 450,000 participants), there are 805,426 markers in the released genotype data. These arrays were augmented by imputation of approximately 90 million genetic variants from the Haplotype Reference Consortium and the UK 10K projects. Details of the array design, sample processing, and stringent quality control have been previously described. we further excluded participants with sex mismatch, heterozygosity rate outliers, and missing genotypes, with excess relatives and non-European ancestry (non self-report British white or Caucasians ascertained by genetic principal
components analysis), as well as prevalent RA participants (ICD-9 codes: 714; ICD-10 codes: M05 and M06; Self report RA:1464 ), resulting in a final cohort of 424,282 individuals (including 4897 incident RA cases)

Quantification and statistical analysis

**Enrichment analysis in medication categories**

We used the prioritized gene list as the primary input to conduct a sequence of Fisher’s exact tests concerning ATC codes, aiming to assess the enrichment of drug-target genes within specific codes. The drug-target database utilized in this analysis was sourced from GREP,\textsuperscript{28} curated from two prominent drug repositories, Drug Bank and TTD. The Anatomical Therapeutic Chemical (ATC) classification system serves as a pivotal framework for categorizing drugs, where approved medications can belong to multiple ATC codes. We selected the ATC code L as RA relevant medication category: anti-tumor and immune modulator, characterized by the highest count of approved drugs tailored for RA treatment in this ATC code medication category.

**Pathway enrichment analysis**

We conducted one-sided hypergeometric tests utilizing the "ReactomePA" R package to perform enrichment analysis on the prioritized genes within Reactome pathways.\textsuperscript{32} Following the default settings, we considered the pathways as significant if \( p \)-values < 0.05 after being adjusted by the Benjamini-Hochberg method.
**Observational analysis at population level**

We used Cox proportional hazard model to calculate hazard ratios and 95% confidence intervals of incident RA associated with BRD2 genotype (rs206783). Model adjusted for age, sex, assessment centers, genotyping arrays and the first 10 genetic PC. The time scale of the cox model is based on the follow-up time, this was defined from the date of initial recruitment to the incident RA, date of death, loss of follow-up, or end of follow-up (31 October 2022), whichever came first.

**Data and code availability**

The RA GWAS summary statistics and whole exome gene burden tests are publicly available at https://www.ebi.ac.uk/gwas/summary-statistics. Plasma pQTL summary statistics are obtained from deCODE study at https://www.decode.com/summarydata/.

The genetic and phenotypic UK Biobank data are available on application to the UK Biobank to any researcher worldwide (www.ukbiobank.ac.uk).

FUMA (https://fuma.ctglab.nl/), PheWAS (https://azphewas.com/) and the Pi and coloc R packages.

Supplementary Information

Document S1. Figures S1 ~ S6

Table S1  Enrichment of prioritized drug-target genes in the ATC codes, Related to Figure 1

Table S2  Genes nominated by the omnibus gene prioritization, Related to Figure 1

Table S3  Top 5% significant genes associated with RA from gene burden test results computed on WES data from the UK Biobank

Table S4  Enrichment of prioritized drug-target genes in the ATC codes, based on prioritized genes from Table S3

Table S5  Associations of 1881 plasma proteins with the risk of rheumatoid arthritis using SMR & HEIDI methods, related to Figure 2

Table S6  Drug candidates for drug-target proteins with MR evidence, Related to Figure 2

Table S7  Pathways significantly enriched in the genes prioritized by three approaches, related to Figure 4

Table S8  Phenotypes associated with BRD2 at variant level by PheWAS analysis, related to Figure 5
References


doi:10.1038/s41588-021-00913-z


doi:10.1038/s41588-021-00978-w


doi:10.1038/s41586-021-04103-z


doi:10.1016/j.xgen.2023.100341

doi:10.1016/j.xgen.2022.100190


doi:10.1371/journal.pcbi.1004219


doi:10.1038/s41588-019-0367-1


doi:10.1038/s41590-019-0378-1


doi:10.1002/art.3862


63. Wang Q, Yang C, Gelernter J, Zhao H. Pervasive pleiotropy between


Study aim: To discover novel targets and repurpose drugs for rheumatoid arthritis (RA) by integrating genomics-driven approaches

Methods and Results

Main data resource

RA GWAS summary statistic: 22,350 RA cases, 74,823 controls, both European ancestry

Whole exome gene burden tests: 6100 RA cases, 379,952 controls

Plasma protein QTLs: 4,907 aptamers in 35,559 Icelanders

Drug target database: (https://therapeutics.discoverytarget.org)
L: ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS

<table>
<thead>
<tr>
<th>Drug Name (E.g.)</th>
<th>Indication (E.g.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>belatacept</td>
<td>Kidney transplant rejection</td>
</tr>
<tr>
<td>palbociclib</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>denileukin</td>
<td>Organ transplant rejection</td>
</tr>
<tr>
<td>sunitinib</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>carfilzomib</td>
<td>Multiple myeloma</td>
</tr>
<tr>
<td>carfilzomib</td>
<td>Multiple myeloma</td>
</tr>
<tr>
<td>fingolimod</td>
<td>Multiple sclerosis</td>
</tr>
</tbody>
</table>

ATC Code

A: ALIMENTARY TRACT AND METABOLISM
B: BLOOD AND BLOOD FORMING ORGANS
C: CARDIOVASCULAR SYSTEM
D: DERMATOLOGICALS
E: ENDOCRINE SYSTEM
F: EYES AND ORBIT
G: GENITO URINARY SYSTEM AND SEX HORMONES
H: SYSTEMIC HORMONAL PREPARATIONS, EXCL. SEX HORMONES AND INSULINS
J: ANTINFECTIVES FOR SYSTEMIC USE
K: KINETICS
L: ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS
M: MUSCULO-SKELETAL SYSTEM
N: NERVOUS SYSTEM
P: ANTIPARASITIC PRODUCTS, INSECTICIDES AND REPELLENTS
R: RESPIRATORY SYSTEM
S: SENSORY ORGANS
V: VARIOUS

For L from ATC code

OR for L from ATC code

Method
- DEPICT
- MAGMA
- Omnibus
- Pi
- PoPS
- TWAS
Omnibus gene prioritization

Exome gene burden test

Proteome-wide Mendelian randomization

Transcriptional regulation by RUNX3

JAK-STAT signaling pathway

PI3K/AKT Signaling

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway

Transcriptional regulation by RUNX3

JAK-STAT Signaling Pathway

PI3K/AKT Signaling Pathway