International Journal of Public Health Science (IJPHS)
Vol. 15, No. 3, September 2026, pp. 848~859
ISSN: 2252-8806, DOI: 10.11591/ijphs.v15i3.27002

Integrative functional annotation of rheumatoid arthritis risk genes using a multi-database bioinformatics approach

Muhammad Nuh, Lolita
Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received Jan 27, 2026
Revised Mar 7, 2026
Accepted Jul 21, 2026

Keywords:

Bioinformatics
eQTL
KEGG
Missense SNPs mutation
Rheumatoid arthritis

ABSTRACT

Rheumatoid arthritis (RA) is an autoimmune disease involving the interaction of genetic and immunological factors. Genome-wide association studies (GWAS) have identified many RA risk loci, but the biological mechanisms linking genetic variation to disease pathogenesis are not yet fully understood. This study aims to prioritize RA candidate genes through a multi-database bioinformatics approach. SNPs significantly associated with RA were obtained from the GWAS catalog, followed by linkage disequilibrium (LD) screening and functional annotation to identify missense variants. The data were integrated with cis-expression quantitative trait loci (cis-eQTL) information, gene ontology (GO) annotation, and Kyoto Encyclopedia of Genes and Genomes (KEGG) molecular pathway mapping. Genes with a total score ≥2 were classified as RA risk genes. A total of 3.145 RA-significant SNPs were identified, of which 58 were missense variants that could potentially affect protein function. The integration of cis-eQTL and functional annotation resulted in a number of candidate genes with the highest scores (score = 4), where *TYK2*, *IL23R*, and *IRAK1* were identified as priority RA genes in the main immune pathways, namely JAK–STAT signaling, IL-23/Th17 axis, and Toll-like receptor–NF-κB signaling. These findings demonstrate that this multi-database-based bioinformatics approach successfully identifies RA candidate genes with strong biological relevance.

Corresponding Author:

Lolita
Faculty of Pharmacy, Universitas Ahmad Dahlan
Yogyakarta, Indonesia
Email: lolita@pharm.uad.ac.id

1. INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disease that primarily affects the joints and is characterized by persistent synovial inflammation, cartilage destruction, bone erosion, and long-term functional disability [1]. Globally, the prevalence of RA is approximately 17.6 million cases, with a higher incidence in women than in men, and the burden of disease increases with age [2]. In Indonesia, the prevalence of RA was reported by Riskesdas to be 11.9% based on medical diagnosis and increased to 24.7% based on clinical symptoms. This confirms the significant impact of RA on the public health system [3].

RA is a complex disease with a strong genetic contribution. Evidence from family and twin studies shows that genetic factors contribute substantially to the risk of RA, with heritability estimated to be around 50-60%. This indicates that genetic variation plays an important role in determining an individual's susceptibility to RA [4]-[6]. Genetic factors not only influence RA risk but also play a role in various stages of pathogenesis, including impaired immune tolerance, activation of autoreactive T and B cells, and regulation of proinflammatory cytokine production. Furthermore, genetic variation also contributes to the heterogeneity of

RA clinical manifestations and differences in the dominant immunological pathways activated in each individual [7], [8]. Therefore, the identification and characterization of RA risk genes is a fundamental step in bridging genetic association findings with an understanding of the molecular mechanisms of the disease, as well as revealing the biological pathways underlying chronic autoimmune inflammation.

Various genomic studies, particularly genome-wide association studies (GWAS), have identified thousands of single nucleotide polymorphisms (SNPs) associated with an increased risk of RA. Previous studies have reported that 23 non-synonymous SNPs directly affect protein function, including rs2476601 in the *PTPN22* gene, rs5029941 and rs2230926 in the *TNFAIP3* gene, and rs34536443 in the *TYK2* gene [9]. Additionally, other genetic variations involved in T cell activation and regulation, such as rs181758100 in the *TNFSF4* gene, rs3181096 in the *CD28* gene, and rs11571319 in the *CTLA4* gene [10]. Meanwhile, SNPs rs16917496 in the 3-UTR region of the *SET8* gene has been reported to act as a predictor of RA risk through the regulation of reactive oxygen species (ROS) and interleukin-10 (IL-10) levels [11]. However, most of the SNPs identified through GWAS are located in non-coding regions or have functional effects that are not yet fully understood, does the relationship between genetic variation and the molecular mechanisms of RA is still not entirely clear.

In addition to genetic variation, the pathogenesis of RA is also influenced by the regulation of gene expression through (cis-eQTL) mechanisms, changes in protein structure and function due to missense mutations, and the involvement of genes in biological processes and molecular signaling pathways. These factors play an important role in regulating the immune response and chronic inflammation that characterize RA [12], [13]. However, most studies still evaluate these aspects separately, whether from the perspective of missense mutations, gene expression regulation, or functional annotation of GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) molecular pathway analysis. As a result, approaches that integrate these various layers of genetic information into a comprehensive analytical framework remain limited [1], [14]. This gap limits a comprehensive understanding of the functional role of RA risk genes and hinders the identification of molecular pathways and more precise therapeutic targets. Advances in bioinformatics enable the integration of various public databases such as GWAS catalog, HaploReg, WebGestalt, UniProt, and KEGG to systematically and multidimensionally annotate the functional risk genes. This multi-database approach allows for more comprehensive mapping between genetic variation and biological function, and has the potential to reveal key immunological pathways involved in RA pathogenesis [15], [16].

Despite the growing number of genetic studies on rheumatoid arthritis, several important research gaps remain. Most previous studies have focused on identifying RA-associated loci through GWAS without systematically integrating functional evidence such as missense variants, gene expression regulation, and pathway involvement into a unified prioritization framework [14]. Moreover, existing bioinformatics studies often evaluate these genetic layers independently, limiting their ability to identify genes with consistent biological relevance across multiple functional dimensions [15]. In the Indonesian context, where RA prevalence and disease burden are increasing, and molecular epidemiological data remain scarce, this lack of integrative genetic analysis hampers the translation of genomic findings into public health strategies, including early risk identification and targeted intervention planning [3]. Therefore, an explicit integrative bioinformatics approach that connects GWAS signals with functional annotation and prioritization is needed to bridge this gap and provide molecular insights that are relevant not only for biological understanding but also for public health decision-making in Indonesia.

Based on this background, this study aims to perform integrated functional annotation of RA risk genes using a multi-database bioinformatics approach. This approach combines missense SNPs analysis, cis-eQTL, GO annotation, and KEGG molecular pathway analysis. It is expected to strengthen the understanding of the functional role of RA risk genes and support the development of a scientific framework for more precise molecular target identification.

2. METHOD

This study is a multi-database bioinformatics study that aims to identify and prioritize RA risk genes through integrated functional annotation using public genomic databases. The analysis was conducted by combining information from missense SNPs, cis-eQTLs, GO, and UniProt KB-KEGG molecular pathways to obtain a more comprehensive biological understanding of RA pathogenesis.

2.1. Bioinformatics data sources

This study is based on a secondary analysis utilizing various publicly available bioinformatics data, including the GWAS catalog (https://www.ebi.ac.uk/gwas/), HaploReg version 4.2 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php), WebGestalt (https://www.webgestalt.org/), and UniProt KB-KEGG (https://www.uniprot.org/) (https://www.kegg.jp/), accessed on October 28, 2025 as the main analysis. The process of searching for bioinformatics data sources can be seen in Figure 1.

All genomic coordinates were harmonized to the human reference genome build GRCh37 (hg19) to ensure consistency across GWAS summary statistics, eQTL datasets, and functional annotation resources. The GWAS data analyzed in this study were derived primarily from publicly available studies conducted in European and East Asian populations, reflecting the dominant reference populations in large-scale RA genetic studies. Expression quantitative trait loci (eQTL) data were obtained from publicly available repositories based on immune-relevant tissues and whole blood, which are commonly used in autoimmune disease research.

2.1.1. Data selection rationale

The selection of bioinformatics databases in this study was based on their scientific validity, data curation standards, and relevance to genetic and functional annotation of rheumatoid arthritis. The GWAS catalog was chosen as the primary source of genetic variants because it provides systematically curated, peer-reviewed associations with standardized significance thresholds, ensuring robust and reproducible variant selection. HaploReg v4.2 was selected due to its ability to integrate linkage disequilibrium information with functional annotations, including missense variants and cis-expression quantitative trait loci (cis-eQTLs), which are essential for assessing regulatory and functional effects of genetic variants. WebGestalt was used for gene ontology (GO) analysis because it enables statistically controlled over-representation analysis across biological processes, molecular functions, and cellular components. UniProtKB and KEGG were selected for pathway mapping as they provide manually curated protein and pathway information, allowing reliable interpretation of molecular mechanisms involved in immune and inflammatory signaling. The combined use of these databases enables a complementary and multi-layered assessment of RA risk genes that cannot be achieved using a single data source [15], [17].

2.1.2. GWAS catalog

The GWAS catalog is used as the primary source for identifying all SNPs that have been reported to be associated with RA. SNPs searches are performed using the keyword "rheumatoid arthritis" and filtered based on the standard genomic significance threshold ($p < 5\times10^{-8}$) to ensure that only strong and reliable association variants are analyzed further. This significance threshold is in accordance with the official GWAS catalog guidelines, which are updated regularly and widely used in genomic studies to improve data consistency and reproducibility [17], [18].

2.1.3. HaploReg version 4.2

HaploReg v4.2 is used for functional annotation of SNPs obtained from the GWAS catalog. This database provides information on variant types, including identification of missense SNPs, as well as LD data to assess relationships between variants within a genetic locus. In addition, HaploReg is used to evaluate the potential regulatory role of SNPs through regulatory element annotation, cis-eQTL analysis, and minor allele frequency (MAF) information relevant to understanding variant distribution in populations [19], [20].

2.1.4. WebGestalt

WebGestalt was used for GO annotation analysis of identified candidate genes. GO analysis covers three main categories, namely biological process, cellular component, and molecular function, to identify the dominant biological functions of RA risk genes. The over-representation analysis (ORA) method is applied to determine significantly enriched GO categories, enabling functional interpretation of genes in the context of RA pathogenesis [21], [22].

2.1.5. UniProt KB-KEGG

Uniprot KB is used to obtain curated protein information, including protein function, structural domains, and protein involvement in specific biological processes. Data from UniProt KB is then mapped into molecular pathways using KEGG to identify biological and immunological pathways relevant to RA. This pathway analysis aims to link risk genes to specific molecular mechanisms underlying inflammation and immune dysregulation in RA [23]-[25]. Integrative multi-database bioinformatics can be seen in Figure 1.

2.2. Bioinformatics analysis

2.2.1. Data collection and quality control of variants

Genetic variants associated with RA were collected from the GWAS catalog using the keyword "rheumatoid arthritis." All SNPs were filtered based on the standard genomic significance threshold ($p < 5\times10^{-8}$) to ensure that only variants with strong associations were analyzed further. The application of this threshold aims to minimize the possibility of false positives due to large-scale genomic association testing. This process resulted in a total of 3,145 SNPs that met the quality criteria for analysis in the next stage. All gene symbols were verified and standardized using the HUGO Gene Nomenclature Committee (HGNC)

database to ensure consistent and accurate gene identification. Deprecated symbols and aliases were cross-checked and updated prior to downstream analyses [26], [27].

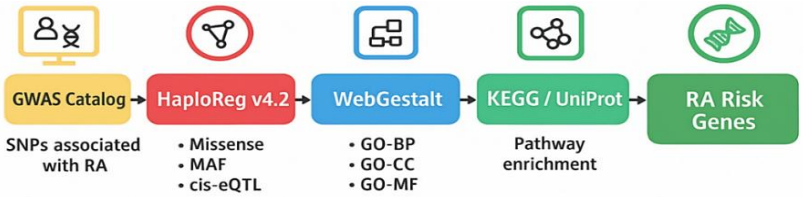

Figure 1. Integrative multi-database bioinformatics analysis of rheumatoid arthritis risk genes

2.2.2. Linkage disequilibrium (LD) screening

LD screening was performed using HaploReg v4.2 with a correlation coefficient threshold of $r^2 > 0.8$. This step aims to reduce variant redundancy originating from the same genetic association signal in a single locus. By eliminating highly correlated SNPs, the analysis focuses on the most representative variants in each LD block. This strategy is commonly used in the fine-mapping process to prioritize variants with higher potential biological relevance [18], [28].

2.2.3. Functional annotation: missense mutations and minor allele frequency (MAF)

SNPs resulting from LD screening were analyzed using HaploReg v4.2 to identify missense variations, namely nucleotide variants that cause amino acid changes and have the potential to directly impact protein structure and function. From this analysis, 58 missense SNPs were prioritized as candidates with direct functional impact. In addition, each SNPs was annotated with MAF information in Asian, African, European, and American populations to evaluate cross-population allele distribution and assess variant relevance across populations [29]-[31].

2.2.4. Gene expression regulation annotation (cis-eQTL)

Cis-eQTL analysis was performed using HaploReg v4.2 to evaluate the potential influence of SNPs on gene expression regulation around target locations. SNPs showing an association with changes in gene expression levels were recorded as variants with potential regulatory functions. This stage is important for linking genetic variation with transcriptional regulatory mechanisms that contribute to RA pathogenesis [32], [33].

2.2.5. Gene ontology (GO) analysis

GO analysis was performed using WebGestalt with the ORA approach. The analysis covered three main categories, namely biological process, cellular component, and molecular function, to identify the dominant biological functions of candidate genes. The significance threshold was set at a false discovery rate (FDR) < 0.05 to control for errors due to multiple testing. The results of the GO analysis were used to systematically interpret the biological roles of RA risk genes [34].

2.2.6. Molecular pathway mapping KEGG pathway

The next candidate genes are mapped into the KEGG molecular pathway using UniProtKB. This step aims to identify biological and immunological pathways involved in RA pathogenesis. The identified pathways include major immune signaling pathways such as JAK–STAT, NF-κB, toll-like receptor, and inflammatory cytokine pathways known to play a role in autoimmune responses [20], [35].

2.2.7. Multi-database integration and RA risk gene prioritization

All annotation results are integrated using the multilayer gene prioritization approach. Each gene is assigned a score based on evidence support from multiple annotation layers, namely the presence of missense mutations (score 1), cis-eQTL evidence (score 1), involvement in KEGG pathways (score 1), and GO analysis significance (score 1). Genes with a total score ≥ 2 are classified as priority RA risk genes that have higher biological relevance and potentially play an important role in RA pathogenesis. The integrative scoring framework was designed to prioritize genes supported by convergent functional evidence across multiple annotation layers rather than relying on a single data source. Each gene received one score per functional layer, including missense variant annotation, cis-eQTL evidence, GO enrichment, and KEGG pathway involvement. This equal-weight scoring approach was chosen to reduce analytical bias and improve interpretability. Genes with higher cumulative scores were considered higher-priority candidates due to consistent multi-layer biological support [36], [37]. The multi-database bioinformatics can be seen in Figure 2.

2.2.8. Potential bias and limitations of secondary data

This study is subject to several potential biases inherent to the use of secondary data from public databases. First, GWAS data are influenced by population representation bias, as the majority of published studies are derived from populations of European ancestry, which may limit the generalizability of findings to other populations, including Indonesia. Second, functional annotations such as cis-eQTLs and pathway assignments depend on database coverage and tissue availability, potentially underrepresenting disease-relevant tissues in rheumatoid arthritis. Third, linkage disequilibrium patterns may vary across populations, which can affect variant prioritization. To mitigate these limitations, this study applied stringent genome-wide significance thresholds, linkage disequilibrium filtering, and multi-layer functional integration to prioritize genes supported by consistent evidence across independent annotation layers. Nevertheless, the findings should be interpreted as hypothesis-generating and require experimental validation and population-specific studies to confirm biological and public health relevance [27], [28].

2.2.9. Statistical analysis

To assess the robustness of the prioritization results, a sensitivity analysis was conducted by evaluating gene rankings under alternative scoring scenarios, including partial exclusion of individual annotation layers. Core prioritized genes remained consistently ranked among the highest-scoring candidates, indicating that the results were not driven by a single data source but reflected stable integrative signals. The statistical analysis can be seen in Figure 2.

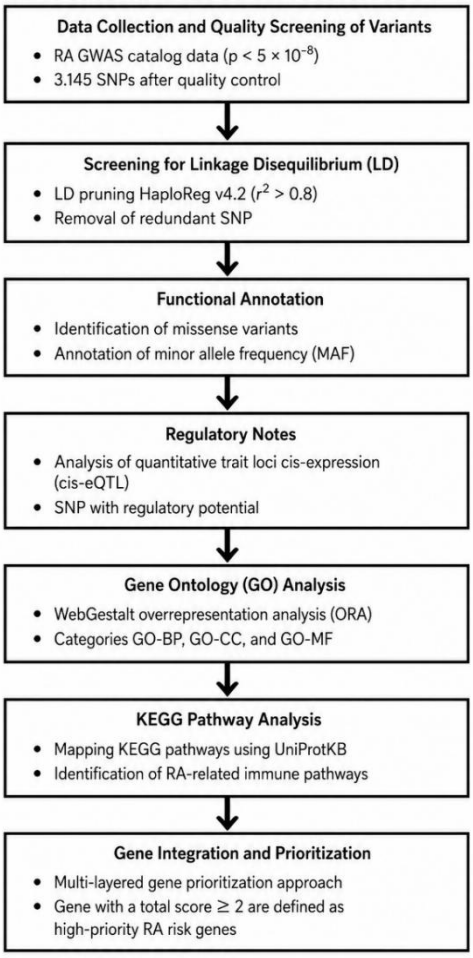

Figure 2. Multi-database integrative bioinformatics workflow for functional annotation and prioritization of RA risk genes

3. RESULTS AND DISCUSSION

Bioinformatics analysis was performed to identify and prioritize gene candidates associated with RA based on genetic evidence characteristics and functional annotation. The analysis pipeline included screening

SNPs from the GWAS catalog, identifying missense variants, integrating cis-eQTL data, GO annotation, and mapping molecular pathways using KEGG. These data layers were integrated into a layered prioritization framework to identify RA risk genes with consistent functional evidence. Gene nomenclature and prioritization scores were applied consistently across the main text, tables, and figures to ensure coherence in data presentation.

3.1. Results

3.1.1. SNPs screening and missense variant identification

A total of 3,145 SNPs significantly associated with RA ($p < 5\times10^{-8}$) were identified from the GWAS Catalog. Next, LD screening was performed using HaploReg v4.2 with a threshold of $r^2 > 0.8$ to reduce genetic signal redundancy. This step produced a number of representative variants that were then functionally annotated. Functional annotation identified 58 missense SNPs that could potentially cause amino acid changes in target proteins. These missense variants were distributed across various candidate genes commonly associated with immune function, inflammation, and cellular response regulation. All genes containing these missense SNPs were compiled as a basis for further analysis and systematically summarized in Table 1.

3.1.2. Integration of cis-eQTL and candidate gene identification

Integration of variant data with cis-eQTL was performed using HaploReg v4.2 to strengthen the biological relevance of candidate genes. This approach enabled the identification of SNPs with the potential to alter proteins and gene regulators. The results showed that 52 candidate genes achieved an integrative score ≥ 2, indicating support from multiple layers of functional evidence. A small number of genes showed lower scores due to limited annotation support. The distribution of scores and functional annotations are presented in detail in Table 1.

Although a substantial number of genes were identified through missense variant analysis alone, integration with cis-eQTL data revealed that genes supported by both structural variation and transcriptional regulation evidence achieved higher prioritization scores. This finding indicates that variants influencing both protein structure and gene expression have greater biological significance in RA. Compared to genes supported by only a single annotation layer, these variants are more consistently associated with RA-related biological relevance.

3.1.3. Functional classification based on GO

Gene ontology analysis was performed to classify the functions of candidate biological genes into three main categories, namely biological processes, cellular components, and molecular functions. The annotation results show that some candidate genes are involved in processes related to immune response regulation, signal transduction, and intercellular communication. At this stage, GO annotation is used as a descriptive functional classification without mechanistic interpretation. All GO annotation results are summarized as part of the integrative gene score and are shown in Table 1.

3.1.4. Molecular pathway mapping using KEGG

Molecular pathway mapping was performed by linking candidate genes to the KEGG database via UniProtKB. This analysis identified the involvement of a number of genes in molecular pathways related to cytokine signaling, innate immune response, immune cell activation, and antigen presentation. The KEGG mapping results in this section are presented descriptively as part of the integrative score assessment, without further biological interpretation. Complete information regarding the involvement of candidate genes in KEGG pathways is presented in Table 1.

3.1.5. Identification of priority RA risk genes

Based on the results of combining all parameter analyses, candidate genes were ranked to identify genes with the most consistent involvement. The ranking results showed that *TYK2*, *IRAK1*, and *IL23R* had the highest integrative scores (score = 4), representing involvement in all major analysis parameters. Based on the integrative scoring results, *TYK2*, *IL23R*, and *IRAK1* were identified as priority RA risk genes with the highest total scores. These genes showed consistent support across all functional annotation layers, including missense variants, cis-eQTL evidence, GO annotation, and KEGG pathway involvement. A summary of the functional roles, molecular mechanisms, and immune-related pathways associated with these prioritized genes is presented in Table 2.

3.2. Discussion

This study applied a multi-database-based bioinformatics approach to prioritize RA candidate genes based on a combination of missense SNPs, cis-eQTL regulation, GO annotation, and KEGG molecular pathway mapping. This approach enables the identification of genes that are not only genetically associated but also show consistent functional involvement in key biological pathways of the immune system. Based on the highest integrative score, *TYK2*, *IL23R*, and *IRAK1* were identified as priority RA risk genes and became the main focus of the discussion.

Table 1. Multilayer functional annotation and integrative scoring of candidate RA risk genes

No	Gene	Missense	cis-eQTL	Biological process	Cellular component	Molecular function	KEGG	Total score
1	*HLA-F*	1	1	0	0	0	1	3
2	*ZNF679*	1	0	0	0	0	0	1
3	*SMAD3*	1	0	0	0	0	1	2
4	*REV3L*	1	1	0	0	0	0	2
5	*HLA-C*	1	1	0	0	0	1	3
6	*POM121L2*	1	1	0	0	0	1	3
7	*TRAF3IP2-AS1*	1	1	0	0	0	1	3
8	*UBD*	1	0	0	0	0	0	1
9	*IRAK1*	1	1	1	0	0	1	4
10	*IFIH1*	1	0	0	0	0	1	2
11	*FCGR2A*	1	1	0	0	0	1	3
12	*ICOSLG*	1	1	0	0	0	1	3
13	*STATUS4*	1	0	0	0	0	0	1
14	*RIN3*	1	1	0	0	0	0	2
15	*IL6R*	1	1	0	0	0	1	3
16	*ZPBP2*	1	1	0	0	0	0	2
17	*GSDMB*	1	1	0	0	0	0	2
18	*RTKN2*	1	1	0	0	0	0	2
19	*DNASE1L3*	1	1	0	0	0	0	2
20	*PRKCQ*	1	0	1	0	0	1	3
21	*PTPN22*	1	1	1	0	0	0	3
22	*TLR4*	1	1	0	0	0	1	3
23	*ICAM3*	1	1	0	0	0	1	3
24	*AIFM2*	1	0	0	0	0	1	2
25	*FUT2*	1	1	0	0	0	1	3
26	*NOS2*	1	0	0	0	0	1	2
27	*WDFY4*	1	1	0	0	0	0	2
28	*SH2B3*	1	1	0	0	0	1	3
29	*NOD2*	1	1	0	0	0	1	3
30	*HLA-DQB1*	1	1	0	0	0	1	3
31	*C6ORFL5*	1	1	0	0	0	1	3
32	*TAP2*	1	1	0	0	0	1	3
33	*BSN*	1	1	0	0	0	0	2
34	*MST1*	1	1	0	0	0	1	3
35	*HLA-DOB*	1	1	0	0	0	1	3
36	*HLA-DMA*	1	1	0	0	0	1	3
37	*ERAP1*	1	1	0	0	0	0	2
38	*DCLRE1B*	1	1	0	0	0	0	2
39	*FAM8A1*	1	0	0	0	0	0	1
40	*TYK2*	1	1	1	0	0	1	4
41	*IL23R*	1	1	1	0	0	1	4
42	*GPR35*	1	1	0	0	0	1	3
43	*CCHCR1*	1	1	0	0	0	0	2
44	*GCKR*	1	1	0	0	0	0	2
45	*HSPA1L*	1	1	0	0	0	1	3
46	*CUTA*	1	1	0	0	0	0	2
47	*TCTE1*	1	0	0	0	0	0	1
48	*NFKBIE*	1	1	0	0	0	1	3
49	*AARS2*	1	0	0	0	0	1	2
50	*BTNL2*	1	1	0	0	0	0	2
51	*TNXB*	1	1	0	0	0	1	3
52	*NCF2*	1	0	0	0	0	1	2
53	*CD6*	1	1	0	0	0	0	2
54	*PSORS1C2*	1	1	0	0	0	0	2
55	*USP8*	1	0	0	0	0	1	2
56	*ATF6B*	1	0	0	0	0	1	2
57	*MEFV*	1	0	0	0	0	0	1
58	*ZFP57*	1	1	0	0	0	0	2

Table 2. Functional summary of prioritized RA risk genes

No	Gene	Gene function	Molecular mechanism	Association with the immune system
1	*IRAK1*	IL-1 or TLR signal transduction	Kinases that activate the NF-κB and MAPK pathways	Regulation of innate immune response
2	*TYK2*	Cytokine signal transduction	JAK family kinases that mediate IL-12, IL-23, and interferon signaling	Regulation of immune and inflammatory responses
3	*IL23R*	Cytokine receptor	IL-23 receptor subunit that activates the JAK-STAT pathway	Differentiation and maintenance of Th17 cells

TYK2, as one of the genes with the highest integrative score, occupies a central component in the cytokine signaling network relevant to RA. *TYK2* is a member of the Janus kinase (JAK) family that plays a role in mediating the signal transduction of various proinflammatory cytokines, including interleukin-12 (IL-12), interleukin-23 (IL-23), and interferon. Activation of these pathways contributes to the differentiation and activation of immune cells, as well as maintaining the chronic inflammation that characterizes RA [38]-[40]. This integrative result is consistent with prior clinical and experimental research, including evidence from international cohorts wherein *TYK2* variants have been associated with RA susceptibility and functional impact on immune pathways. In addition, studies in East Asian populations have similarly reported associations between *TYK2* polymorphisms and RA risk, supporting the relevance of *TYK2* across diverse genetic backgrounds. This convergence of genetic and functional annotation evidence reinforces the effectiveness of JAK pathway modulation in suppressing inflammatory activity and improving clinical manifestations of RA, as demonstrated in clinical trials of JAK inhibitors that target *TYK2*-mediated signaling [41], [42]. Thus, the identification of *TYK2* in the present study not only aligns with the international literature but also underscores its potential translational relevance to population-specific RA research and therapeutic strategies.

IL23R shows consistent involvement across all layers of JAK–STAT signaling pathway analysis. *IL23R* is the major receptor subunit for interleukin-23 (IL-23), a cytokine that plays an important role in the differentiation, expansion, and maintenance of T helper 17 (Th17) cells. Activation of the IL-23/Th17 axis contributes to the production of proinflammatory cytokines known to play a role in amplifying synovial inflammation and joint tissue damage in RA [43]-[45]. The prioritization of *IL23R* in this integrative bioinformatics analysis is consistent with previous international studies demonstrating that dysregulation of the IL-23/Th17 pathway represents one of the dominant immune mechanisms sustaining chronic synovial inflammation in RA [46]. In addition, recent genomic and immunological studies have reported significant associations between *IL23R* variants and RA susceptibility, further supporting the role of *IL23R*-mediated signaling in disease pathogenesis across different populations [47]. Thus, the identification of *IL23R* as a priority gene in this study reinforces the central contribution of the Th17 axis to RA and supports its relevance as a potential target for immunomodulatory and precision-based therapeutic strategies.

Meanwhile, *IRAK1* is consistently identified in all layers of analysis and mapped in the Toll-like receptor (TLR)–NF-κB signaling pathway. *IRAK1* is a key kinase in TLR receptor and interleukin-1 (IL-1) receptor signal transduction, which plays a central role in the activation of innate immune responses. Activation of this pathway triggers downstream complex phosphorylation and activation of the NF-κB transcription factor, which in turn increases the expression of various proinflammatory mediators. Dysregulation of the TLR–NF-κB pathway has been shown to contribute to the amplification of chronic inflammation and tissue damage in RA. Previous studies have shown that genetic variations and increased *IRAK1* activity are associated with increased inflammatory responses and disease severity in RA [48], [49]. The prioritization of *IRAK1* in this integrative analysis is consistent with previous studies reporting associations between *IRAK1* genetic variants and heightened inflammatory responses as well as increased disease severity in RA, including findings from population-based research and molecular investigations [50], [51]. Thus, the identification of *IRAK1* in this study underscores the importance of innate immune signaling in RA pathogenesis and supports *IRAK1*-mediated pathways as relevant targets for inflammation-focused and immunomodulatory therapeutic strategies.

The enriched GO terms related to cytokine-mediated signaling, immune activation, and inflammatory responses, together with KEGG pathways such as JAK–STAT signaling, Toll-like receptor signaling, and Th17 cell differentiation, reflect key immunopathological mechanisms underlying RA. These pathways are known to regulate both innate and adaptive immune responses, contributing to persistent synovial inflammation, immune cell infiltration, and joint damage. Collectively, *TYK2*, *IL23R*, and *IRAK1* represent three major immune axes that interact in the pathogenesis of RA, namely cytokine signaling (*TYK2*), the IL-23/Th17 pathway (*IL23R*), and primary inflammatory activation via TLR–NF-κB (*IRAK1*). The integration of these three pathways (Figures 3 and 4) reflects the complexity of immune regulation in RA and provides a relevant molecular framework for the development of more targeted and precise biomarkers and therapeutic strategies.

Although this study successfully identified priority RA candidate genes through a multi-database bioinformatics approach, it has several limitations. The analysis was based entirely on secondary data from public databases, does the results depend on the quality and coverage of the available data. In addition, the bioinformatics approach used has not been complemented by experimental validation, so the causal relationship between genetic variation and RA pathogenesis cannot yet be confirmed. Therefore, further research involving functional validation and clinical studies is needed to confirm the biological and clinical relevance of the identified priority genes.

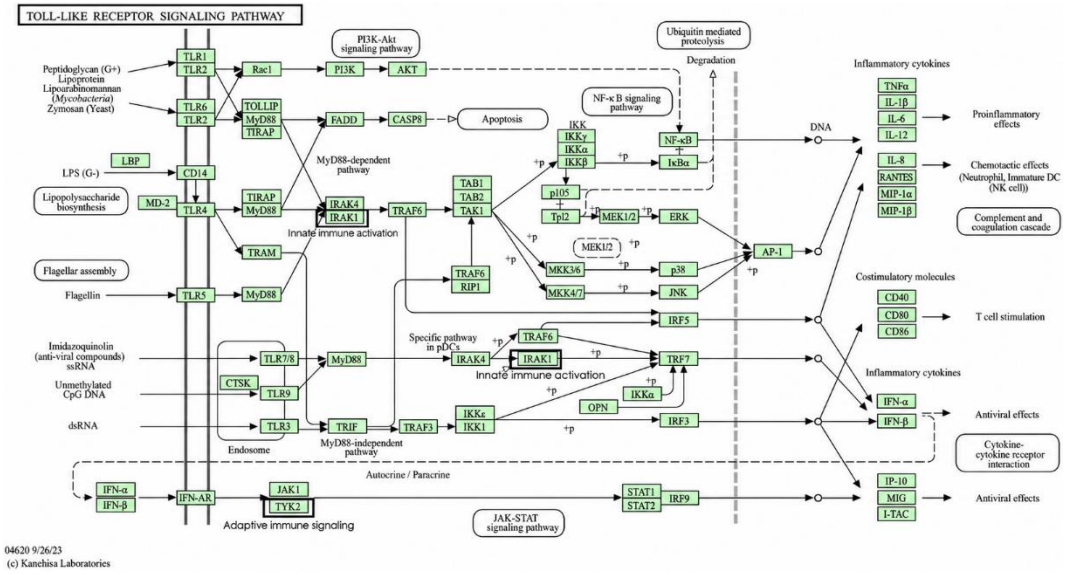

Figure 3. Integrative visualization toll-like receptor signaling pathway of *IRAK1*, *TYK2* involvement in key immunological signaling pathways associated with RA

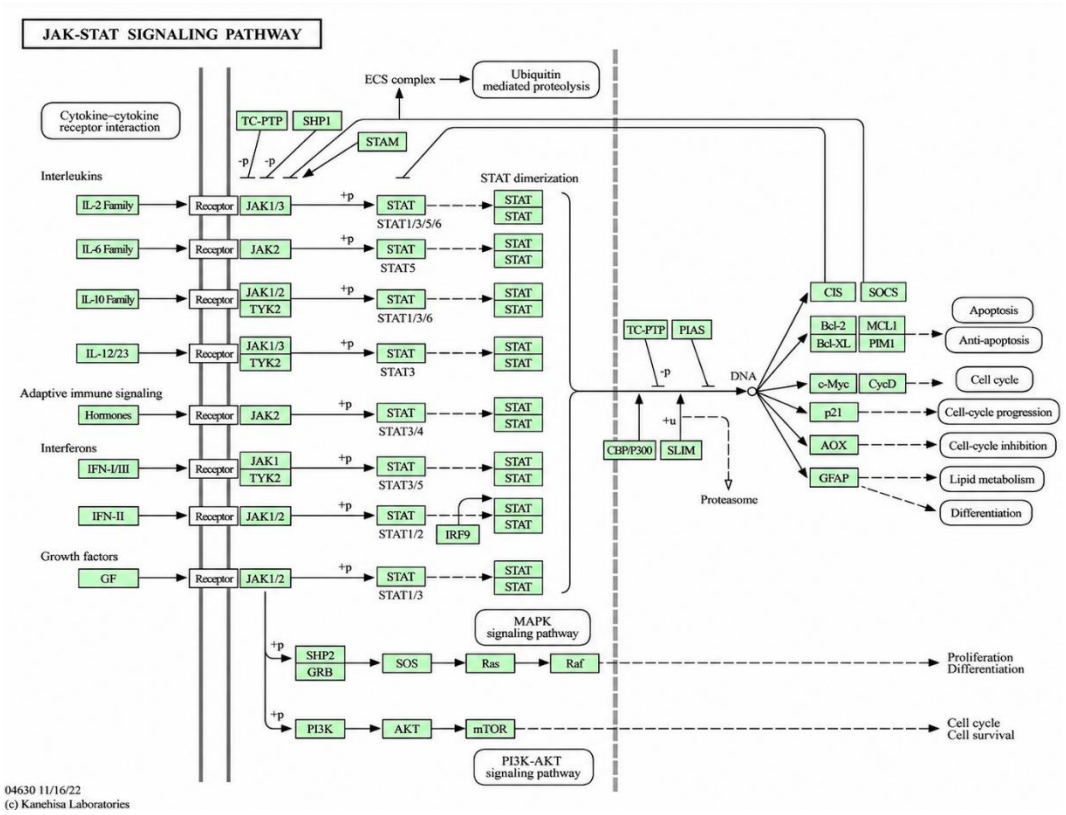

Figure 4. Integrative visualization of the JAK-STAT signaling pathway of *IL23R* involvement in key immunological signaling pathways associated with RA

The findings discussed above are directly supported by integrative genomic evidence, whereas the potential clinical and translational implications should be interpreted cautiously and considered hypothesis-generating. Importantly, we have now elaborated the clinical and public health implications of these findings. From a clinical perspective, the identification of *TYK2*, *IL23R*, and *IRAK1* as prioritized genes may contribute to improved risk stratification, early detection, and the development of targeted therapeutic strategies, particularly in

the context of precision medicine approaches for RA management. From a public health perspective, understanding the genetic and molecular mechanisms underlying RA susceptibility may support the development of population-based screening strategies, improve early intervention efforts, and ultimately reduce disease burden, disability, and long-term healthcare costs associated with RA. These findings also provide a valuable foundation for future translational research aimed at improving preventive and personalized treatment strategies.

4. CONCLUSION

This study demonstrates that a multi-database-based bioinformatics approach is effective in identifying candidate genes with consistent genetic and functional support in RA. The integration of missense SNPs, cis-eQTL, GO annotation, and KEGG pathway mapping data places *TYK2*, *IL23R*, and *IRAK1* as priority genes involved in key immune pathways that play a role in RA pathogenesis. These findings provide more comprehensive molecular insights into the inflammatory mechanisms of RA and support the potential of these three genes as biomarkers and therapeutic targets. However, experimental validation and further clinical studies are still needed to confirm the biological relevance and clinical implications of the identified priority genes.

ACKNOWLEDGMENTS

The author would like to thank Universitas Ahmad Dahlan for its institutional support, as well as the supervising lecturers who provided guidance and input, particularly in the field of bioinformatics, does that this research could run smoothly.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial or notfor-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal adopts the Contributor Roles Taxonomy (CRediT) to ensure transparency and clarity of individual author contributions. Each author has made substantial contributions to the work and meets the authorship criteria.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muhammad Nuh	✓	✓			✓				✓	✓	✓		✓	✓
Lolita		✓		✓	✓	✓		✓	✓			✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

This study did not involve direct interaction with human participants. All analyses were conducted using anonymized and publicly available datasets. Therefore, informed consent was not required.

ETHICAL APPROVAL

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was not required because the study used anonymized, publicly available data and did not involve direct interaction with human participants.

DATA AVAILABILITY

The main data analyzed in this study were obtained from public databases, namely GWAS Catalog, HaploReg v.4.2, WebGestalt, and KEGG-UniPro**t**. The processed data resulting from this analysis are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. M. Birga, L. Ren, H. Luo, Y. Zhang, and J. Huang, "Prediction of new risk genes and potential drugs for rheumatoid arthritis from multiomics data," *Computational and Mathematical Methods in Medicine*, 2022, doi: 10.1155/2022/6783659.

[2] R. J. Black, M. Cross, L. M. Haile, G. T. Culbreth, and J. D. Steinmetz, "Global, regional, and national burden of rheumatoid arthritis, 1990–2020, and projections to 2050: a systematic analysis of the global burden of disease study 2021," *The Lancet Rheumatology*, vol. 5, 2023, doi: 10.1016/S2665-9913(23)00211-4.

[3] Herlina, S. Rahutami, and N. S. Murni, "Hubungan pola makan dan aktivitas fisik dengan kejadian rheumatoid arthritis pada lansia Puskemas Tebing Gerinting Kabupaten Ogan Ilir tahun 2024," *Jurnal Kesehatan Tambusai*, vol. 5, no. 2, 2024, doi: 10.31004/jkt.v5i2.28232.

[4] J. K. Muruganantham and R. Veerabathiran, "The role of genetic variants in rheumatoid arthritis pathophysiology and their impact on disease mechanisms," *Journal of Health Science and Medical Research*, vol. 43, no. 1, 2025, doi: 10.31584/jhsmr.20241061.

[5] N. Tchitchek *et al.*, "Deep immunophenotyping reveals that autoimmune and autoinflammatory disorders are spread along two immunological axes capturing disease inflammation levels and types," *Annals of The Rheumatic Diseases*, vol. 83, 2024, doi: 10.1136/ard-2023-225179.

[6] I. Brown, "The role of genetics in rheumatoid arthritis: insights into disease pathogenesis and treatment," *Journal of Clinical Case Reports*, vol. 15, no. 02, 2025.

[7] J. Zhao, S. Guo, S. J. Schrodi, and D. He, "Molecular and cellular heterogeneity in rheumatoid arthritis: mechanisms and clinical implication," *Frontiers in Immunology*, vol. 12, 2021, doi: 10.3389/fimmu.2021.790122.

[8] M. Masoumi *et al.*, "The genetic puzzle of rheumatoid arthritis: causes, progression, and treatment," *Biochemistry and Biophysics Reports*, vol. 43, 2025, doi: 10.1016/j.bbrep.2025.102148.

[9] M. Akhtar *et al.*, "Characterization of rheumatoid arthritis risk-associated SNPs and identification of novel therapeutic sites using an in-silico approach," *Biology*, vol. 10, no. 6, 2021, doi: 10.3390/biology10060501.

[10] D.-P. Chen, Y.-H. Wen, W.-T. Lin, F.-P. Hsu, and K.-H. Yu, "Exploration of the association between the single-nucleotide polymorphism of co-stimulatory system and rheumatoid arthritis," *Frontiers in Immunology*, vol. 14, 2023, doi: 10.3389/fimmu.2023.1123832.

[11] C. Peng, Y. Zhao, X. Zhang, J. Zhang, Z. Sha, and S. Zhang, "Polymorphisms in microrna binding site of set8 regulate the risk of rheumatoid arthritis," *Experimental and Therapeutic Medicine*, vol. 25, no. 244, 2023, doi: 10.3892/etm.2023.11943.

[12] N. M. Sushanti, W. Adikusuma, A. N. Puspitaningrum, A. R. Afief, and L. M. Irham, "Drug repurposing for rheumatoid arthritis through the utilization of genetic variation data," *Medical Sains*, vol. 8, no. 2, 2023, doi: 10.37874/ms.v8i2.720.

[13] L. Chen, J. Zhao, and Q. Meng, "From genetic variants to therapeutic targets: insights into understanding rheumatoid arthritis," *Frontiers in Immunology*, vol. 16, 2025, doi: 10.3389/fimmu.2025.1556971.

[14] L. Wu *et al.*, "Potential mechanisms and drugs prediction of rheumatoid arthritis and primary Sjögren's syndrome: a public databases-based study," *PLOS ONE*, vol. 19, no. 2, 2024, doi: 10.1371/journal.pone.0298447.

[15] S. Jahangir, P. John, A. Bhatti, M. M. Aslam, and M. J. Peffers, "Data integration for big data analytics to identify the gaps in rheumatoid arthritis genomics in a post-GWAS era," *International Conference on Advance Computing and Innovative Technologies in Engineering*, vol. 2, 2022, doi: 10.1109/ICACITE53722.2022.9823601.

[16] A. Torres *et al.*, "Epigenetic regulation of nutrient transporters in rheumatoid arthritis fibroblast-like synoviocytes," *Arthritis Rheymatol*, vol. 74, no. 7, 2022, doi: 0.1002/art.42077.

[117] S. Qiu, A. Munir, S. I. Malik, S. Khan, and A. Hassan, "Identification of differentially expressed genes and pathways crosstalk analysis in rheumatoid and osteoarthritis using next-generation sequencing and protein-protein networks," *Saudi Journal of Biological Sciences*, vol. 28, 2021.

[18] N. M. Sushanti *et al.*, "Mapping rheumatoid arthritis susceptibility through integrative bioinformatics and genomics," *Journal of Pharmaceutical Science*, vol. 20, no. 1, pp. 15–23, 2023, doi: 10.12928/mf.v20i1.24912.

[19] R. Mungmunpuntipantip and V. Wiwanitkit, "Bioinformatics approach for searching for natural products in vector-borne disease management," *North Clin Istanb*, vol. 11, no. 1, pp. 171–176, 2024, doi: 10.14744/nci.2023.87523.

[20] G. Anuraga, D. Anurogo, F. Fitriani, H. B. Rochmanto, and Z. B. Widya, "Integrative bioinformatics and statistical approaches for identifying prognostic biomarkers and therapeutic targets in breast cancer," *Eigen Mathematics Journal*, vol. 8, no. 1, 2025, doi: 10.29303/emj.v8i1.277.

[21] J. Xu *et al.*, "Identification of candidate genes related tosynovial macrophages in rheumatoid arthritis bybioinformatics analysis," *International Journal of General Medicine*, vol. 14, 2021, doi: 10.2147/IJGM.S333512.

[22] X. He *et al.*, "Identification and validation of potential hub genes in rheumatoid arthritis by bioinformatics analysis," *American Journal of Translational Research*, vol. 14, no. 9, 2022.

[23] Z. Li *et al.*, "Drug discovery in rheumatoid arthritis with joint effusion identified by text mining and biomedical databases," *Annals of Palliative Medicine*, vol. 10, no. 5, 2021, doi: 0.21037/apm-20-2631b.

[24] L. Najafzadeh, M. Mahmoudi, M. Ebadi, M. D. Shasaltaneh, and A. Masoudinejad, "Identification of key genes and pathways for differential diagnosis between rheumatoid arthritis and osteoarthritis based on systems biology approaches," *Rheumatology Research Journal*, vol. 6, no. 1, 2021, doi: 10.1097/MD.0000000000010997.

[25] F. Yu, G. Hu, L. Li, Boyu, and R. Liu, "Identification of key candidate genes and biological pathways in the synovial tissue of patients with rheumatoid arthritis," *Experimental and Therapeutic Medicine*, vol. 23, 2023, doi: 10.3892/etm.2022.11295.

[26] Y. Xiong, B.-B. Mi, M.-F. Liu, H. Xue, Q.-P. Wu, and G.-H. Liu, "Bioinformatics analysis and identification of genes and molecular pathways involved in synovial inflammation in rheumatoid arthritis," *Medical Science Monitor*, vol. 25, 2019, doi: 10.12659/MSM.915451.

[27] Z. Chen, M. Boehnke, X. Wen, and B. Mukherjee, "Revisiting the genome-wide significance threshold for common variant GWAS," *Genes, Genomes, Genetics*, vol. 11, no. 2, 2021, doi: 10.1093/g3journal/jkaa056.

[28] Q. S. Wang and H. Huang, "Methods for statistical fine-mapping and their applications to auto-immune diseases," *Seminars in Immunopathology*, vol. 44, no. 1, 2022, doi: 10.1007/s00281-021-00902-8.

[29] K. Wang, H. Luo, L. Liu, H. Gao, Y. Song, and D. Li, "Blockade of il-1 family cytokines in the treatment of rheumatoid arthritis," *Frontiers in Pharmacology*, vol. 16, 2025, doi: 10.3389/fphar.2025.1577628.

[30] Y. Wu *et al.*, "Genome-wide fine-mapping improves identification of causal variants," *medRxiv*, vol. 27, 2024, doi: 10.1101/2024.07.18.24310667.

[31] Z. Mousavi *et al.*, "Prioritization of causal genes from genome-wide association studies by Bayesian data integration across loci," *PLOS Computational Biology*, vol. 21, no. 1, 2025, doi: 10.1371/journal.pcbi.1012725.

[32] K. Luo, Y. Zhong, Y. Guo, J. Nie, Y. Xu, and H. Zhou, "Integrated bioinformatics analysis and experimental validation reveals hub genes of rheumatoid arthritis," *Experimental and Therapeutic Medicine*, vol. 26, no. 4, 2023, doi: 10.3892/etm.2023.12179.
[33] F. Jiang *et al.*, "A landscape of gene expression regulation for synovium in arthritis," *Nature Communications*, vol. 15, no. 1409, 2024, doi: 10.1038/s41467-024-45652-x.
[34] L. Milchram *et al.*, "Functional analysis of autoantibody signatures in rheumatoid arthritis," *Molecules*, vol. 27, no. 1452, 2022, doi: 10.3390/molecules27041452.
[35] F. Xu *et al.*, "Detection of common pathogenesis of rheumatoid arthritis and atherosclerosis via microarray data analysis," *Heliyon*, vol. 10, no. 8, 2024, doi: 10.1016/j.heliyon.2024.e28029.
[36] H. Huang, X. Dong, K. Mao, W. Pan, B. Nie, and L. Jiang, "Identification of key candidate genes and pathways in rheumatoid arthritis and osteoarthritis by integrated bioinformatical analysis," *Frontiers in Genetics*, vol. 14, 2023, doi: 10.3389/fgene.2023.1083615.
[37] Y. Liu, S. Fan, and S. Meng, "Identification of the candidate genes of diagnosing rheumatoid arthritis using the single-cell sequencing technology and t cell subclusters analysis of patients with rheumatoid arthritis," *The Archives of Rheumatology*, vol. 38, no. 1, 2023, doi: 10.46497/ArchRheumatol.2022.9573.
[38] A. F. Izati, K. K. Wong, and C. H. C. Maraina, "IL-23/il-17 axis in the pathogenesis and treatment of systemic lupus erythematosus and rheumatoid arthritis," *Malaysian Journal Pathology*, vol. 42, no. 3, 2020.
[39] Y. Peng, B. Chen, X. Sheng, and Y. Qian, "Polymorphisms in irf5 and TYK2 genes are associated with rheumatoid arthritis in a Chinese Han population," *Medical Science*, vol. 27, 2021, doi: 10.12659/MSM.928455.
[40] S. Yuan *et al.*, "Mendelian randomization and clinical trial evidence supports TYK2 inhibition as a therapeutic target for autoimmune diseases," *The Lancet*, vol. 89, 2023, doi: 10.1016/j.ebiom.2023.104488.
[41] L. Rusiñol and L. Puig, "TYK2 targeting in immune-mediated inflammatory diseases," *International Journal of Molecular Sciences*, vol. 24, 2023, doi: 10.3390/ijms24043391.
[42] B. Yang, L. Chu, F. Feng, S. Lu, and C. Xue, "Association of tyrosine kinase 2 polymorphisms with susceptibility to microscopic polyangiitis in a guangxi population," *PeerJ*, vol. 12, 2024, doi: 10.7717/peerj.18735.
[43] E. Soysal, F. Ulutaş, E. Tepeli, S. Kaymaz, and V. Çobankara, "IL-23R gene polymorphisms in rheumatoid arthritis," *Rheumatology International*, vol. 42, no. 3, 2022, doi: 10.1007/s00296-021-04881-9.
[44] M. Abbasifard, M. R. Mirzaei, M. Abade, and Z. Bagheri-Hosseinabadi, "Interleukin 23 receptor gene polymorphisms and their role in the inflammatory status of rheumatoid arthritis patients in an Iranian population," *International Journal of Rheumatic Diseases*, vol. 26, no. 2, 2023, doi: 10.1111/1756-185X.14479.
[45] J. Du, X. Wang, G. Tan, Z. Liang, Z. Zhang, and H. Yu, "The association between genetic polymorphisms of interleukin 23 receptor gene and the risk of rheumatoid arthritis: an updated meta-analysis," *Clinical Immunology*, vol. 210, 2020, doi: 10.1016/j.clim.2019.108250.
[46] B. Yucel, C. Sumer, I. Gok, M. Karkucak, E. Alemdaroglu, and F. Ucar, "Associations between cytokine gene polymorphisms and rheumatoid arthritis in Turkish population," *Northern clinics of Istanbul*, vol. 7, no. 6, 2020, doi: 10.14744/nci.2020.70845.
[47] T. Jahan, A. A. Saleh, and S. Anwar, "Association of cytokine IL-17, IL-4, IL-6, and IL-12 gene polymorphisms in rheumatoid arthritis patients in a tertiary care hospital in Bangladesh," *International Journal of Rheumatology*, vol. 2, 2024, doi: 10.1155/2024/3728179.
[48] N. Hosseini, M. T. Tahoori, A. Mohammadzadeh, H. Z. Jaliani, M. B. Sani, and H. S. Salehabadi, "IRAK1 gene polymorphism in rheumatoid arthritis," *A Journal of Molecular and Cellular Immunology*, vol. 50, no. 1, 2021, doi: 10.1080/08820139.2020.1764028.
[49] H. Yan *et al.*, "Associations of IRAK1 gene polymorphisms and mRNA expression with NMOSD risk in the northern Chinese Han population," *Frontiers Neurology*, vol. 12, 2021, doi: 0.3389/fneur.2021.661791.
[50] E. A. Al-Saffar and B. Q. Al-Saadi, "Study the association of IRAK1gene polymorphism and some immunological markers with the risk of rheumatoid arthritis incidence in sample of Iraqi patients," *Iraqi Journal of Biotechnology*, vol. 21, no. 2, 2022.
[51] H. Cen *et al.*, "Associations between genetic polymorphisms within transporter genes and clinical response to methotrexate in Chinese rheumatoid arthritis patients: a pilot study," *Pharmacogenomics and Personalized Medicine*, vol. 15, 2022, doi: 10.2147/PGPM.S350417.

BIOGRAPHIES OF AUTHORS

Muhammad Nuh is a researcher at the Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. He obtained his professional degree in Pharmacy from Sultan Agung Islamic University, Semarang, in 2024. Currently, he is pursuing a master's degree in pharmacy at Universitas Ahmad Dahlan. His current research focuses on clinical and social-behavioral pharmacy, particularly in relation to musculoskeletal disorders and bone health. He can be contacted at email: 2408045013@webmail.uad.ac.id.

Lolita is an associate professor of Ahmad Dahlan University, Yogyakarta, Indonesia. In 2005, she received a Bachelor of Pharmacy from the Faculty of Pharmacy, Ahmad Dahlan University, Yogyakarta, and in 2013, she obtained a Master of Science in Pharmacy from Gadjah Mada University. Since 2013, she has been a member of the teaching staff at the Faculty of Pharmacy, Universitas Ahmad Dahlan. During her academic career, she has taught various courses including pharmacoepidemiology, pharmacotherapy and medicinal chemistry, pharmacy practice, health behavior and outcomes, as well as pharmaceutical skills and applications. Her current research focuses on pharmacoepidemiology, clinical and social–behavioral pharmacy, particularly in relation to infectious diseases and renal disorders. She can be contacted at email: lolita@pharm.uad.ac.id.