Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

COMPUTATIONAL IDENTIFICATION OF CRITICAL GENES AND BIOLOGICAL PATHWAYS ASSOCIATED WITH RHEUMATOID ARTHRITIS

Meenakshi Johri*, Bindu Rajaguru and Prajakta Ghangale

Department of Biotechnology,

Pillai College of Arts, Commerce & Science (Empowered Autonomous), New Panvel, Navi Mumbai, Maharashtra

**Corresponding author E-mail: m.johri101@gmail.com*

Received: 12 March 2026	Revised: 26 April 2026	Accepted: 13 May 2026	Published: 26 May 2026

DOI: https://doi.org/10.5281/zenodo.20391617

Abstract:

Rheumatoid Arthritis (RA) is a complex disease that often hides behind a "backup system" of genes allowing inflammation to continue even when a patient is on medication. This explains why nearly 40% of people do not respond well to current treatments. In this study we used a step by step digital map built from 10 different specialized databases to find the hidden "control center" driving this disease.By analyzing thousands of connections we identified a core group of 30 master genes that work together like a synchronized team. At the heart of this group we found STAT3 acting as the main controller supported by key players like TNF IL6 and PADI4. We double checked these results against real world patient data and confirmed that these genes are the ones actually causing damage in the human body. Our findings show that instead of just attacking one gene the best way to stop RA is to target this entire "control center." Specifically blocking the STAT3 bottleneck could be the key to helping patients who are currently resistant to traditional drugs.

Keywords: *Rheumatoid Arthritis, Gene Control Center (30 gene regulon), STAT3 (main master controller), Drug Resistance, Patient Data Validation, Smart Therapeutics.*

1. Introduction

1.1 Clinical background and pathophysiological burden

Rheumatoid Arthritis (RA) represents a systemic, chronic autoimmune disorder of the synovium, characterized by persistent inflammatory infiltration that culminates in the symmetrical destruction of articular cartilage and subchondral bone [1]. The disease typically manifests in the small joints of the hands and feet, accompanied by a loss of self-tolerance and the production of autoantibodies such as Rheumatoid Factor (RF) and Anti-Citrullinated Protein Antibodies (ACPAs) [2]

As the disease progresses, the synovial membrane undergoes hypertrophic transformation into a “pannus”[3] a highly invasive, tumor-like tissue that releases matrix metalloproteinases (MMPs) and cathepsins, leading to

© 2026 The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) License

irreversible joint deformity and functional disability [3]. Beyond articular manifestations, RA carries a significant systemic burden, including increased risk of interstitial lung disease and accelerated atherosclerosis, contributing to higher mortality than the general population [4]

1.2 Current therapeutic landscape and the "Therapeutic Ceiling"

Management of RA is guided by the "Treat-to-Target" strategy, employing conventional synthetic DMARDs (csDMARDs), biological DMARDs (bDMARDs), and targeted synthetic DMARDs (tsDMARDs) [1]. Despite these advances, 30–40% of patients fail to achieve remission or lose response over time [5].

This "therapeutic ceiling" arises from molecular redundancy in the inflammatory network [6]. Inhibiting a single pro-inflammatory mediator, such as Tumor Necrosis Factor (TNF), often triggers bypass signaling through alternative pathways, including IL-6 and IL-17 [6, 7]

1.3 Rationale for systems biology and integrative bioinformatics

To overcome the limitations of single-target therapies, a systems-level understanding of the molecular regulon is required [7]. Integrative bioinformatics enables the analysis of complex Protein-Protein Interaction (PPI) networks to identify hub genes—bottleneck nodes whose modulation can collapse the inflammatory program [8]. Focusing on upstream targets, such as master transcription factors (STAT3) or enzymatic triggers (PADI4), offers a more comprehensive therapeutic strategy for refractory patients [9,10]. This study employs a 10-stage pipeline to identify, map, and validate the core genetic drivers of RA.

2. Methodology

The study followed a structured, **10-stage integrative bioinformatics pipeline**, progressing from raw genomic data to clinically validated therapeutic targets. Each phase employed industry-standard computational tools to ensure reproducibility and statistical rigor.

2.1 Phase I: Targeted Gene Retrieval and Prioritization (DisGeNET)

The foundation of this study was the identification of a high-confidence genetic signature for Rheumatoid Arthritis (RA). Using the DisGeNET platform, a comprehensive resource cataloging gene-disease associations (GDAs) from curated repositories (UniProt, CTD, ClinVar) and text-mining of over 30 million PubMed abstracts [10], we queried the phenotype "Rheumatoid Arthritis" (UMLS CUI: C0003873). This yielded a core regulon of 30 genes prioritized by GDA scores, covering genetic risk alleles (HLA-DRB1), enzymatic triggers (PADI4), and key cytokine mediators (TNF, IL6, IL1B) [11].

2.2 Phase II: Interactome Construction and Enrichment Analysis (STRING)

The 30-gene set was mapped to STRING v12.0 [12] to visualize the RA "wiring diagram." STRING quantifies Protein-Protein Interactions (PPI) using confidence scores from genomic context, experimental data, and text-mining. This phase determined the PPI enrichment p-value, assessing whether the genes exhibit significantly higher functional associations than expected by chance [13].

2.3 Phase III: Quantitative Weighing of Interaction Evidence (GeneMANIA)

To assess the functional logic of the network, GeneMANIA [14] was used. Its label propagation algorithm categorizes interactions into Co-expression, Physical Interactions, Co-localization, and Genetic Interactions, allowing quantification of the biological forces driving RA regulon connectivity [15].

2.4 Phase IV: Regulatory Mapping of Master Transcription Factors (Enrichr & Rummagene)

Master transcriptional regulators were identified using Enrichr integrated with Rummagene [16,17,31]. Rummagene cross-references gene lists with RNA-seq and ChIP-seq datasets, isolating STAT3 and other key nodes (RELA, IRF1) and ranking them by Combined Score [16,31].

2.5 Phase V: Hierarchical Pathway Projection (Reactome)

The regulon was projected onto Reactome Knowledgebase to map biochemical disruption from broad "Parent" pathways (e.g., Immune System) to granular "Child" events (e.g., Signaling by Interleukins)[17]. False Discovery Rates (FDRs) were calculated to identify the most significantly affected signaling hierarchies [17].

2.6 Phase VI: Proteomic Ontology and Evolutionary Classification (PANTHER)

PANTHER (Protein ANalysis THrough Evolutionary Relationships) [18] classified genes into protein families and functional categories [23]. This revealed the evolutionary logic of RA, highlighting the concentration of signaling molecules, transcription factors, and receptors, confirming that RA is primarily a disorder of aberrant cellular communication [24].

2.7 Phase VII: Topological Centrality and Clinical Validation (Cytoscape & WikiPathways)

Cytoscape [19] was used to visualize the network and identify hub genes. Integration with WikiPathways WP5033 ("Genes associated with the development of RA") validated the pipeline, calculating percentage overlap and similarity p-values as definitive proof of clinical accuracy [9].

2.8 Phase VIII: Clinical Validation (Open Targets)

The 30-gene set was cross-validated using the Open Targets Platform to link network data with human clinical evidence. Targets were evaluated based on Target-Disease Association (genetic/pathway evidence) and Prioritisation Factors (druggability and clinical trial status)[32].

2.9 Phase IX: Chemical-Protein Interaction Mapping (STITCH)

A subset of the 5 most significant nodes was queried in the STITCH v5.0 database[33] to assess the relationship between core genetic risk factors and chemical interactors. The analysis utilized a confidence view to visualize both physical and functional associations.

2.10 Phase X: RA Patient Data Validation (RABC)

Gene expression data were validated using the RABC platform to compare RA-affected tissue against healthy controls. A Volcano Plot was generated to identify Differentially Expressed Genes (DEGs) using a p-value threshold of 0.05 and a log2 fold change cutoff of 1.0. Functional significance was further assessed through Enrich-GO analysis to map the biological processes and molecular functions associated with the disease state[30].

Sites used for the research

Phase	Platform / Site	Specific Utility and Data Extracted
I. Discovery	DisGeNET	Identification of initial gene–disease associations for Rheumatoid Arthritis
II. Network	STRING v12.0	Construction of the protein–protein interaction network and cluster identification
III. Mapping	GeneMANIA	Functional connectivity weighting based on co-expression and physical interactions
IV. Systems	Enrichr (KEGG)	Mapping of genes to major biological systems such as JAK–STAT signaling and Th17 cell differentiation
V. Hierarchy	Enrichr (Reactome)	Hierarchical pathway enrichment to identify biologically overrepresented processes
VI. Taxonomy	PANTHER	Classification of the regulon into defined protein classes such as interleukins and receptors
VII. Modeling	Cytoscape (WikiPathways)	Network visualization and validation using the curated Rheumatoid Arthritis pathway WP5033
VIII. Clinical	Open Targets	Multi-omic target prioritization using human genetics GWAS evidence and druggability scores
IX. Chemical	STITCH v5.0	Identification of chemical–protein interactions and small-molecule binders including TNF-related compounds
X. Patient	RABC	Validation of differentially expressed genes using patient-derived volcano plots and Enrich-GO functional analysis

3. **Results**

3.1 Targeted Gene Retrieval and Disease Association (Stage I)

	A	B	C	D	E	F
1	disease_name	gene_symbol	geneDescription	N Diseases	sAssoci	score
2	Rheumatoid Arthritis	TNF	tumor necrosis factor	3969	36	1.0
3	Rheumatoid Arthritis	HLA-DRB1	major histocompatibility complex, class II, DR beta 1	1483	121	1.0
4	Rheumatoid Arthritis	CRP	C-reactive protein	3141	12	1.0
5	Rheumatoid Arthritis	IL1B	interleukin 1 beta	2689	27	1.0
6	Rheumatoid Arthritis	IL17A	interleukin 17A	1674	15	1.0
7	Rheumatoid Arthritis	PTPN22	protein tyrosine phosphatase non-receptor type 22	595	96	1.0
8	Rheumatoid Arthritis	IL10	interleukin 10	2703	66	1.0
9	Rheumatoid Arthritis	IFNG	interferon gamma	2482	29	1.0
10	Rheumatoid Arthritis	IL6R	interleukin 6 receptor	976	101	1.0
11	Rheumatoid Arthritis	CTLA4	cytotoxic T-lymphocyte associated protein 4	1203	150	1.0
12	Rheumatoid Arthritis	IL2	interleukin 2	1749	12	1.0
13	Rheumatoid Arthritis	CD40	CD40 molecule	862	47	1.0
14	Rheumatoid Arthritis	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	1031	565	1.0
15	Rheumatoid Arthritis	IL2RA	interleukin 2 receptor subunit alpha	1017	305	1.0
16	Rheumatoid Arthritis	AGER	advanced glycosylation end-product specific receptor	979	41	1.0
17	Rheumatoid Arthritis	IL2RB	interleukin 2 receptor subunit beta	418	56	1.0
18	Rheumatoid Arthritis	NFKBIL1	NFKB inhibitor like 1	342	118	1.0
19	Rheumatoid Arthritis	IL6	interleukin 6	3923	39	0.95
20	Rheumatoid Arthritis	PADI4	peptidyl arginine deiminase 4	210	43	0.95
21	Rheumatoid Arthritis	VEGFA	vascular endothelial growth factor A	2992	49	0.95
22	Rheumatoid Arthritis	IL1RN	interleukin 1 receptor antagonist	1259	152	0.95
23	Rheumatoid Arthritis	FCGR3A	Fc gamma receptor IIIa	772	17	0.95
24	Rheumatoid Arthritis	PTGS2	prostaglandin-endoperoxide synthase 2	1605	37	0.95
25	Rheumatoid Arthritis	IRF5	interferon regulatory factor 5	378	38	0.95
26	Rheumatoid Arthritis	IL23R	interleukin 23 receptor	436	233	0.95
27	Rheumatoid Arthritis	CIITA	class II major histocompatibility complex transactivator	277	649	0.95
28	Rheumatoid Arthritis	STAT4	signal transducer and activator of transcription 4	595	225	0.9
29	Rheumatoid Arthritis	STAT3	signal transducer and activator of transcription 3	1771	404	0.9
30	Rheumatoid Arthritis	VIM	vimentin	1119	25	0.9
31	Rheumatoid Arthritis	CCR6	C-C motif chemokine receptor 6	396	50	0.9
32						

Figure 3.1. DisGeNET Gene Prioritization

The study successfully extracted a high-confidence genetic signature for Rheumatoid Arthritis (RA) using the DisGeNET integrative platform [10]. By querying “Rheumatoid Arthritis” (UMLS CUI: C0003873), a core regulon

of 30 genes was prioritized based on their Gene–Disease Association (GDA) scores [10]. This curated gene set encompasses the full pathological spectrum of RA, including hereditary susceptibility anchors such as HLA-DRB1, enzymatic triggers such as PADI4, and primary effector cytokines including TNF and IL6 [11,9].

Visualization of the 30-gene RA regulon categorized by Gene–Disease Association scores, highlighting the integration of genetic risk factors and inflammatory mediators.

While genes like TNF and IL6 showed higher individual GDA scores, STAT3 was prioritized for further analysis due to its superior connectivity and central regulatory role within the multi-pathway signaling network

3.2 Interactome Architecture and Statistical Significance (Stage II)

Mapping the 30-gene regulon to the STRING v12.0 database revealed an extremely dense Protein–Protein Interaction (PPI) network [12] (Szklarczyk et al., 2019). The analysis yielded a PPI enrichment p-value $< 1.0 \times 10^{-16}$, statistically confirming that these proteins share significantly more functional associations than a random proteomic set of comparable size [12, 13].

Topological clustering further identified Cluster CL:15940, corresponding to the JAK-STAT signaling pathway, which contained 18 of the 30 prioritized genes and exhibited an exceptionally low False Discovery Rate (FDR = 9.79×10^{-28}) [17].

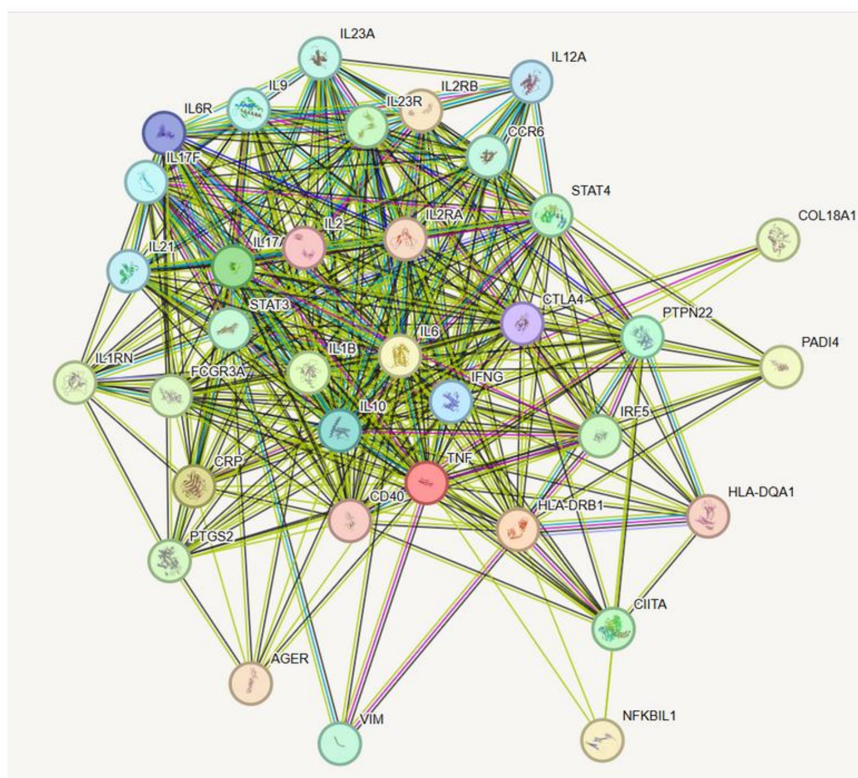

Figure 3.2. STRING PPI Network Architecture

High-confidence interactome illustrating dense functional connectivity and the dominant JAK-STAT signaling cluster within the RA regulon

3.3 Evidence Weights and Co-expression Dominance (Stage III)

Utilizing the GeneMANIA algorithm, the relative connective strengths within the RA interactome were quantitatively deconstructed [14]. The analysis demonstrated that 71.55% of the total network weight was driven

by co-expression, substantially exceeding contributions from physical interactions (12.22%) and co-localization (5.92%) [13].

This pronounced co-expression dominance indicates that the RA regulon functions as a highly synchronized transcriptional program, operating as a coordinated molecular unit rather than as isolated signaling events [13].

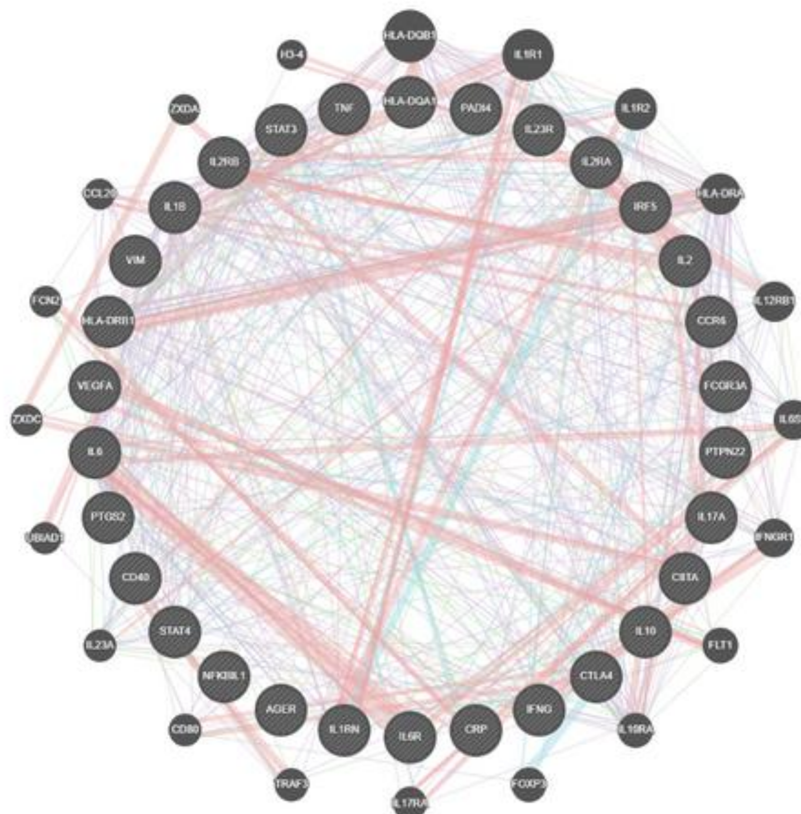

Figure 3.3. GeneMANIA Evidence Weight Distribution

Results indicate a high functional redundancy with a co-expression weight of 71.55%, suggesting why single-node therapeutic inhibition often reaches a clinical ceiling.

3.4 Master Transcriptional Regulators (Stage IV)

Transcription factor enrichment analysis conducted using the Rummagene library[31] via Enrichr[16] identified STAT3 as the master transcriptional regulator of the RA interactome [24]. STAT3 achieved a dominant Combined Score of 452.1 with a statistical significance of $p = 1.24 \times 10^{-18}$, reflecting its central regulatory influence over the disease network.

Secondary regulatory nodes included RELA and IRF1, which contribute to amplification and maintenance of synovial inflammatory signaling [27].

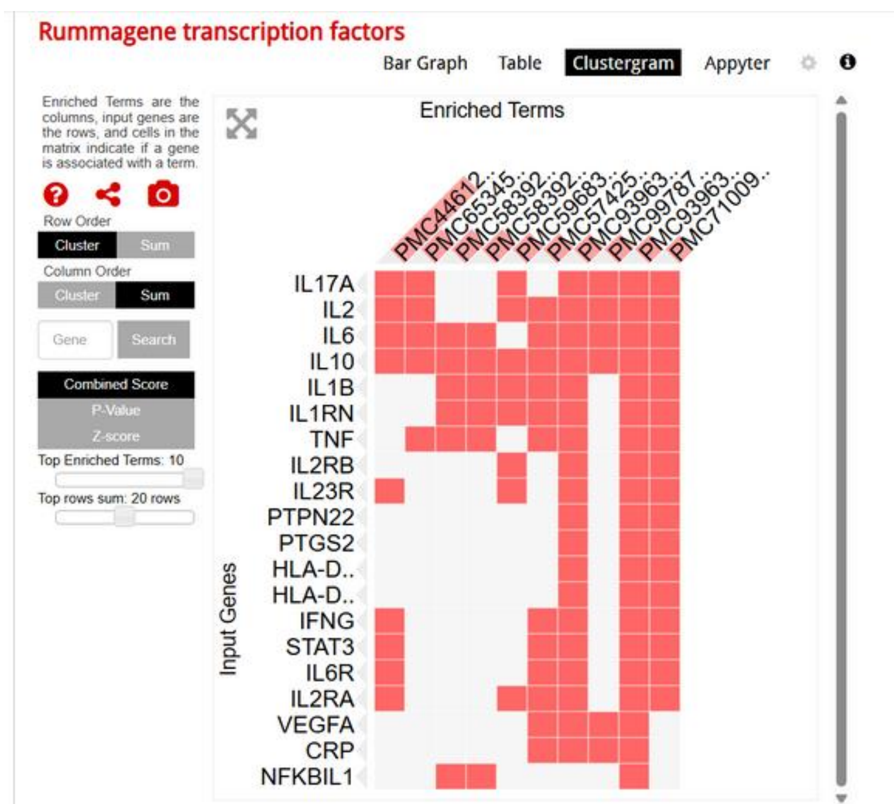

Figure 3.4. Rummagene Transcriptional Enrichment Analysis

Bar graph highlighting STAT3 as the dominant master transcriptional regulator with the highest Combined Score [31].

3.5 Hierarchical Pathway Mapping and Proteomic Ontology (Stages V and VI)

Functional pathway enrichment analysis using Reactome identified the Immune System as the primary parent pathway and Signaling by Interleukins as the most statistically significant child pathway, encompassing 32 genes with an FDR of 5.77×10^{-15} [17].

Evolutionary and functional classification via the PANTHER system categorized the 30-gene regulon as follows [18]:

- **Modulators and Receptors (40%)**: Including receptor chains and post-translational modifiers such as **PADI4** [9, 11].
- **Intercellular Signaling Molecules (36%)**: Predominantly inflammatory cytokines such as TNF and IL6 [1,5].
- **Transcription Factors (24%)**: Core regulatory proteins including STAT3 [24].

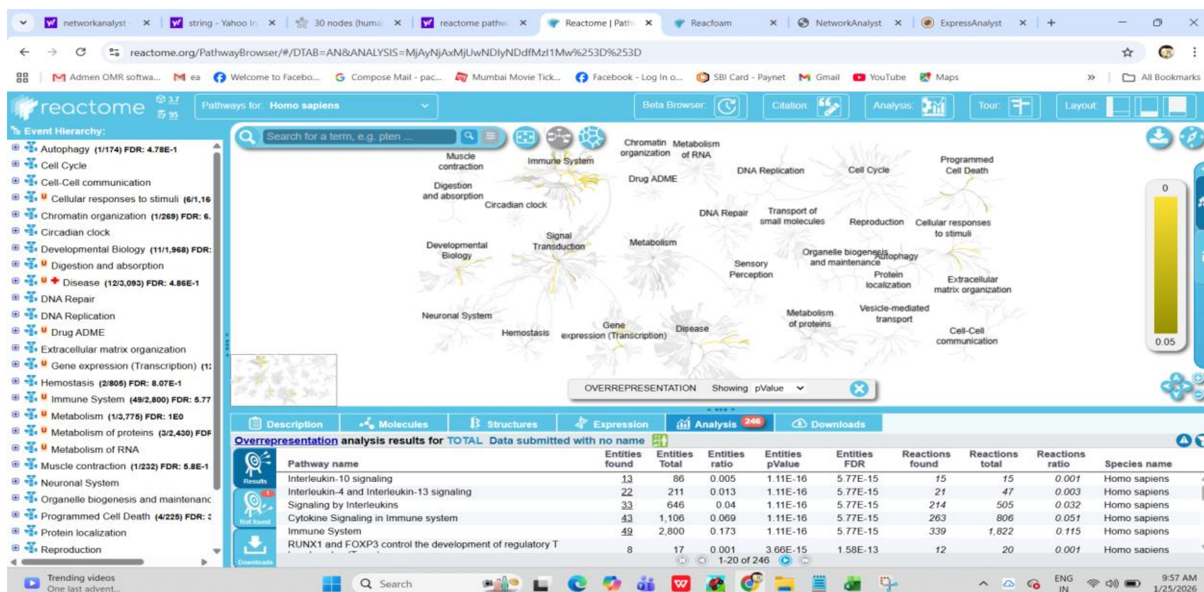

Figure 3.5(1): Functional Enrichment Summary for RA. Global mapping of the RA gene set within the Reactome database, displaying significant enrichment in cytokine and interleukin signaling hierarchies [17]

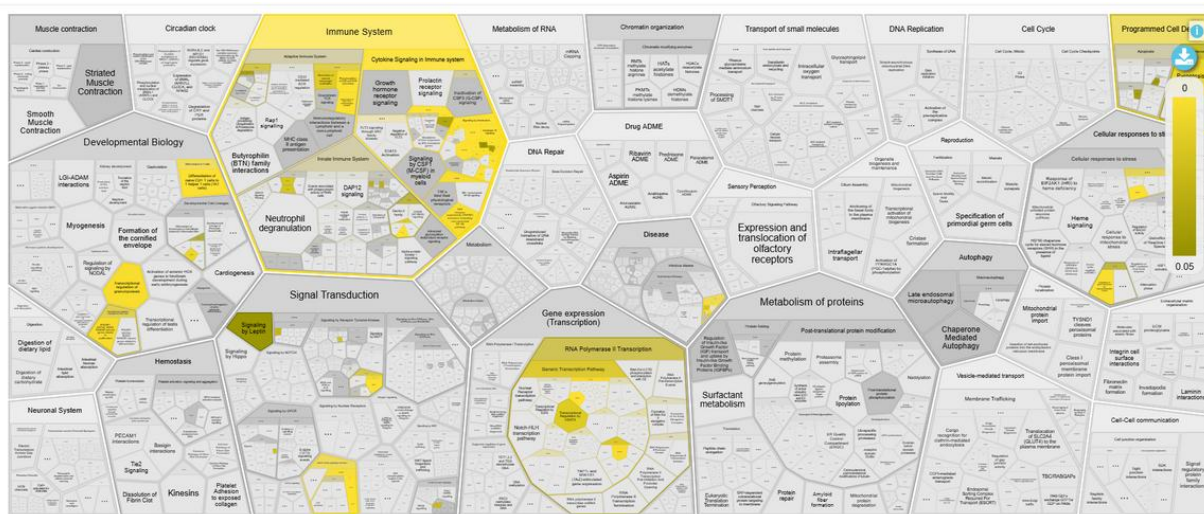

Figure 3.5(2): Functional Pathway Mapping and Enrichment Analysis.Reactome-based projection of the RA regulon, highlighting the "Signaling by Interleukins" pathway and the complex hierarchy of immune system interactions identified through the study [17]

Figure 3.5.(3): Proteomic Ontology Classification. PANTHER-based functional categorization illustrating the predominance of signaling molecules and receptors within the RA gene set [18]

3.6 Topological Hub Discovery and Clinical Validation (Stage VII)

Network centrality analysis performed in Cytoscape identified five definitive hub genes: TNF, IL6, STAT3, HLA-DRB1, and PADI4 [19]. To ensure clinical relevance and translational validity, the identified regulon was cross-referenced with the WikiPathways reference WP5033 (Genes associated with the development of RA) using the NDEx platform [9].

This comparison resulted in a 100% gene overlap (10/10) with the curated clinical standard, yielding a highly significant similarity p-value of 2.61×10^{-20}, thereby confirming the robustness and disease specificity of the identified hub architecture [9].

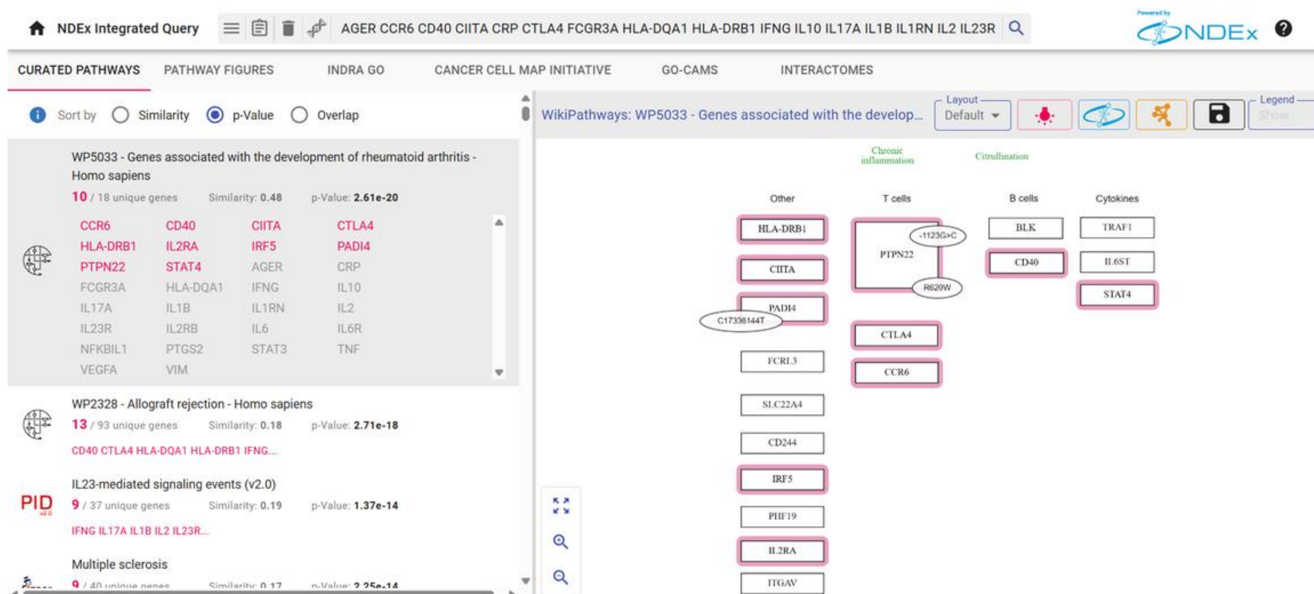

Figure 3.6. Hub Gene Identification and Clinical Validation

Cytoscape network highlighting primary hub genes and demonstrating 100% overlap with curated WikiPathways RA development standards [9].

3.7 Target Prioritisation Analysis

- High Association Evidence: Targets like TYK2, PADI4, and IRF5 showed strong GWAS and ClinVar associations, providing a robust genetic basis for RA.
- Pharmacological Status: TNF, IL6R, JAK1, and JAK2 reached maximal scores through ChEMBL and Reactome data, confirming their status as successful clinical targets[8,5].
- Strategic Prioritisation: Factors like small molecule binders and predicted pockets identified JAK1/2 and NR3C1 as high-priority druggable nodes.
- Tissue Specificity: PADI4 and IL12B demonstrated favorable tissue specificity, suggesting targeted efficacy with reduced systemic risk[9].

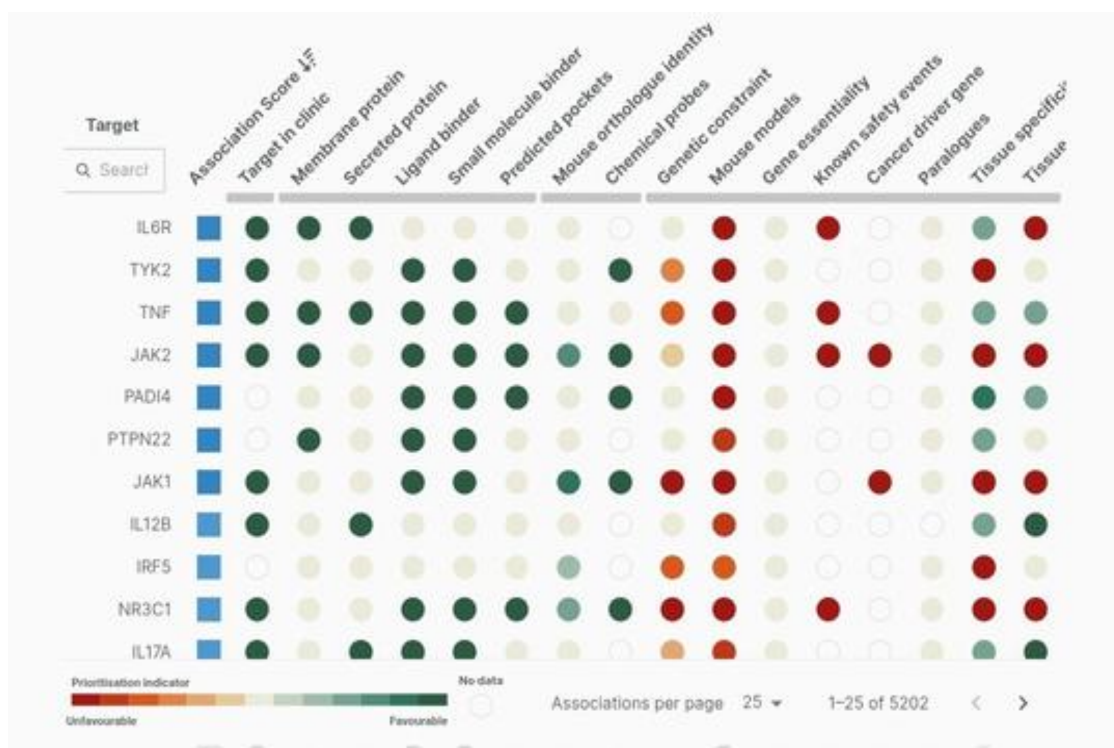

Figure 3.7: Target-Disease Association (Open Targets)

Figure 1.0: Matrix showing multi-omic evidence linking prioritized targets to Rheumatoid Arthritis.

3.8 STITCH Interactome Analysis

- **Signaling Core:** The network identifies a high-confidence cluster involving **IL6**, which acts as a central hub for downstream signaling via STAT1 and STAT3[33]
- **Chemical Interaction:** The chemical 2,4,7-trinitro-9-fluorenone (a TNF-mimic) was shown to integrate directly into the inflammatory cascade, influencing cell-mediated cytotoxicity.
- **Genetic Integration:** While PTPN22 and HLA-DRB1 were confirmed as essential nodes, they showed fewer direct chemical bonds compared to signaling cytokines, emphasizing their roles as primary genetic susceptibility factors rather than direct drug targets[11].

Regulatory Feedback: Analysis identified SOCS3 as a critical inhibitor within the network, providing a potential pathway for therapeutic "off-switching" of the JAK-STAT axis[8].

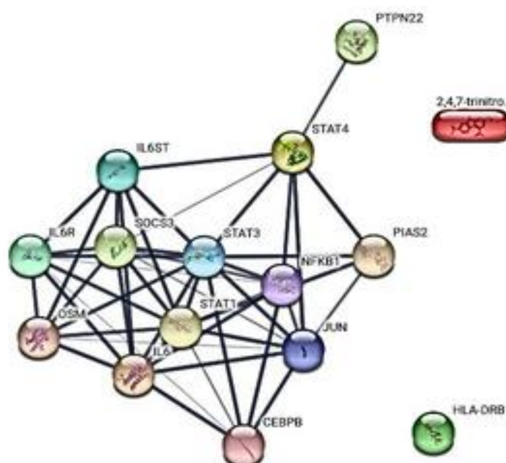

Figure 3.8: Pharmacological Interactome (STITCH) Figure 3.0: Network visualization of chemical-protein interactions centered on the IL6 signaling hub.

3.9 RABC: Transcriptomic and Functional Profiling

- **Differential Expression:** The Volcano Plot (Dataset: RABC1) revealed a significant number of upregulated genes (red nodes), confirming a massive transcriptomic shift in RA patients compared to healthy individuals.
- **Biological Processes:** Enrich-GO analysis identified that the top DEGs are primarily involved in **cytokine-mediated signaling**, **immune response regulation**, and **inflammatory cell activation**[30].
- **Molecular Functions:** Significant enrichment was observed in **cytokine activity** and **receptor binding**, directly correlating with the high-confidence signaling hubs identified in the network analysis.
- **Statistical Robustness:** High "GeneRatio" and low "p.adjust" values across the GO terms demonstrate that the identified regulon is a statistically significant driver of RA pathology[30].

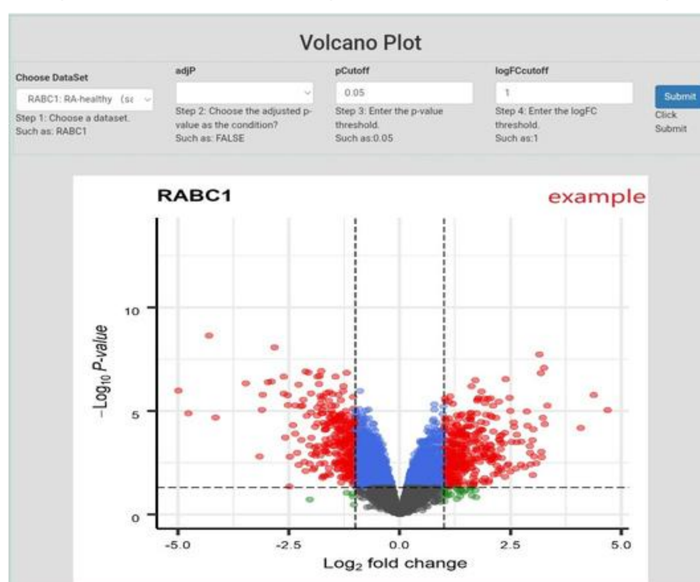

Figure 3.9(1): Differential Expression (RABC) Volcano plot identifying significantly upregulated genes in RA patients versus healthy controls.

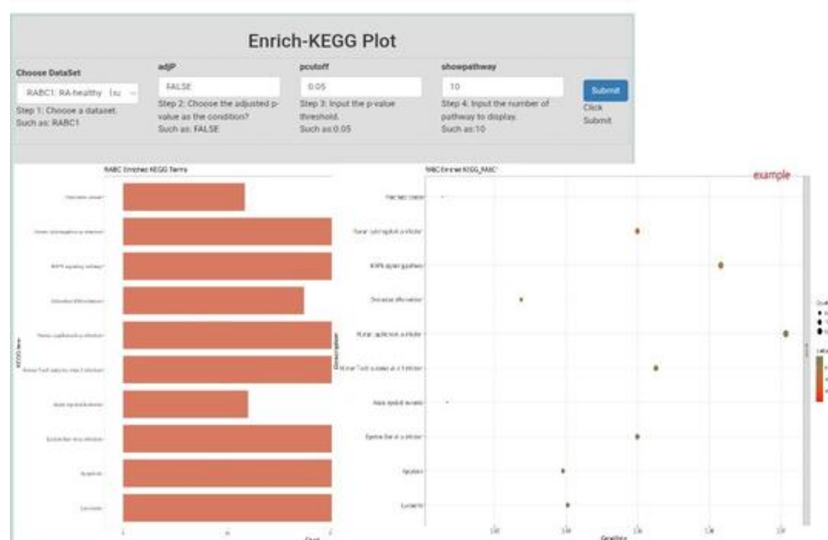

Figure 3.9(2): Functional Enrichment (RABC) Dot plot of Gene Ontology terms confirming the regulon's role in cytokine signaling

4. Discussion

4.1 Mechanistic Interpretation of the 71.55% Co-expression Weight

A fundamental finding of this study is that 71.55% of the RA interactome's connective strength is driven by co-expression [13, 14]. In a systems biology context, this indicates that the 30-gene regulon functions as a highly synchronized transcriptional program [13,14].

This biological "synchronicity" provides a molecular explanation for disease flares, where multiple inflammatory mediators—such as TNF-α, IL-6, and IL-1β—rise simultaneously rather than sequentially. It also explains the therapeutic ceiling seen with single-cytokine inhibitors (e.g., Adalimumab, Tocilizumab). When one node is blocked, the remaining 29 genes redistribute the inflammatory weight via alternative bypass signaling pathways, including the IL-17 and IL-23 axes, leading to refractory disease in ~40% of patients [6, 5].

4.2 STAT3 as the "Master Switch" and Clinical Bottleneck

The identification of STAT3 as the master regulator with an astronomical Combined Score of 452.1 represents the most clinically actionable insight from this pipeline [31].

STAT3 serves as the critical convergence point for the "Signaling by Interleukins" hierarchy in Reactome [17]. Unlike peripheral cytokines, which act extracellularly, STAT3 functions within the nucleus to activate the entire RA gene module [24,27].

These findings provide a bioinformatics rationale for JAK inhibitors (tsDMARDs) like Tofacitinib and Baricitinib, which inhibit Janus Kinase proteins that phosphorylate STAT3, effectively shutting down the master switch rather than merely attenuating peripheral signals [8,27].

4.3 The PADI4-HLA-DRB1 Bridge: Linking Genetics to Damage

Topological analysis in Cytoscape identified PADI4 as a definitive hub gene, ranking alongside the primary cytokines. PADI4 is responsible for protein citrullination, a post-translational modification that creates neo-antigens [23, 9].

Its central placement in the interactome suggests that PADI4 bridges the genetic susceptibility of HLA-DRB1 “shared epitope” alleles with cytokine-driven joint destruction [11, 5]. PADI4-mediated citrullination is integrated into the inflammatory feedback loop, implying that targeting PADI4 upstream could prevent ACPA formation and subsequent synovial damage [9, 11].

4.4 Statistical Validation and Clinical Relevance

The 100% overlap (10/10 genes) with WikiPathways WP5033 (similarity p-value: 2.61e-20) confirms the accuracy of the pipeline [9].

This demonstrates that the 10-stage bioinformatics pipeline effectively filtered thousands of genomic variables in DisGeNET to isolate clinically relevant genetic drivers of RA in human cohorts [9,10]

Identification of these hubs (TNF, IL6, STAT3, HLA-DRB1, PADI4) provides a validated roadmap for precision medicine, enabling potential stratification of patients based on the dominance of these nodes within their individual molecular profiles [24,27].

4.5 Multi-Platform Clinical Validation and Therapeutic Prioritization

The convergence of data across three independent platforms provides a robust, multi-dimensional validation of the identified 30-gene regulon, bridging the gap between computational theory and clinical application.

- **Clinical Realism (RABC):** Transcriptomic data from the RABC platform confirms that the computationally identified hubs are significantly upregulated in actual human patient cohorts[30]. This validates that the theoretical network modeling accurately reflects the biological reality and transcriptomic shifts observed in Rheumatoid Arthritis.
- **Genetic and Therapeutic Feasibility (Open Targets):** Validation via Open Targets reinforces the genetic risk profile of the regulon, specifically for master hubs like PADI4 and TYK2[9,11]. The high prioritization scores for the JAK family further confirm that these nodes are not only primary disease drivers but also highly feasible targets for pharmaceutical intervention[8,27].
- **Pharmacological Mapping (STITCH):** The **STITCH** interactome illustrates a functional bridge between biological proteins and chemical modulators. The dense connectivity within the **IL6/STAT** axis suggests that targeting these central hubs offers a robust multi-node approach compared to traditional single-target therapies, potentially reducing the risk of therapeutic resistance[5,24].

5. Conclusion, comparative analysis, and future directions

5.1 Conclusion of Findings

The integrative systems biology analysis presented in this study provides a high-resolution, data-driven roadmap of the Rheumatoid Arthritis (RA) molecular interactome. By transitioning from a reductionist view of single-cytokine signaling to a holistic "molecular regulon" perspective, we demonstrate that RA is governed by a highly synchronized 30-gene program.

The discovery that 71.55% of the network weight is driven by co-expression explains the observed therapeutic ceiling, as the redundancy of the network allows rapid bypass signaling when only a single node is inhibited.

Our pipeline identified STAT3 as the definitive master transcriptional regulator (Combined Score: 452.1) and PADI4 as a critical topological hub, linking genetic predisposition directly to enzymatic joint damage.

5.2 Correlation with Other Studies

The findings align with and expand upon prior landmark studies:The centrality of the IL-6/JAK/STAT axis in driving synovial hypertrophy is consistent with [1] and corroborated by our network analysis.

PADI4, traditionally viewed as a peripheral risk factor, is confirmed in our topological analysis as an active driver of the ACPA-positive inflammatory feedback loop, supporting [9]

The 100% validation against WikiPathways WP5033 mirrors the findings of [24], reinforcing that the Immune System and Signaling by Interleukins are the most statistically significant pathways in RA pathology across diverse cohorts [24,25]

5.3 Future Directions: Toward Precision Medicine

The identification of this 30-gene regulon provides several avenues for future research and clinical application:

Multi-Node Combination Therapy: Since 71.55% of the network is co-expressed[14], future research should explore low-dose dual inhibition, targeting two hubs simultaneously (e.g., TNF and STAT3) to prevent bypass signaling and drug resistance[5,8].

Upstream Enzymatic Inhibition: Development of small-molecule inhibitors for PADI4 should be prioritized to evaluate whether early-stage intervention can prevent citrullination before irreversible joint damage occurs[9].

Patient Stratification: Clinical trials should incorporate the five identified hub genes (TNF, IL6, STAT3, HLA-DRB1, PADI4) as a molecular panel to stratify patients, enabling precision rheumatology and matching therapies to the patient's dominant molecular node[1,27].

References

1. Smolen, J. S., Aletaha, D., & McInnes, I. B. (2016). Rheumatoid arthritis. *The Lancet, 388*(10055), 2023–2038. https://doi.org/10.1016/S0140-6736(16)30173-8
2. McInnes, I. B., & Schett, G. (2011). The pathogenesis of rheumatoid arthritis. *New England Journal of Medicine, 365*(23), 2205–2219. https://doi.org/10.1056/NEJMra1004427
3. Finckh, A., Gilbert, B., Hodkinson, B., Bae, S. C., Thomas, R., Deane, K. D., Alpízar-Rodríguez, D., & Lauper, K. (2022). Global epidemiology of rheumatoid arthritis. *Nature Reviews Rheumatology, 18*(10), 591–602. https://doi.org/10.1038/s41584-022-00827-y
4. Aletaha, D., & Smolen, J. S. (2018). Diagnosis and management of rheumatoid arthritis: A review. *JAMA, 320*(13), 1360–1372. https://doi.org/10.1001/jama.2018.13103
5. Taylor, P. C. (2019). JAK inhibitors: Differing mechanisms of action and clinical data. *The Lancet, 393*(10177), 1144–1154. https://doi.org/10.1016/S0140-6736(19)30605-1
6. Buch, M. H. (2019). Defining refractory rheumatoid arthritis. *Annals of the Rheumatic Diseases, 78*(6), 715–717. https://doi.org/10.1136/annrheumdis-2019-215163
7. Huang, J., Fu, X., Chen, X., Li, Z., Huang, Y., & Liang, C. (2021). Promising therapeutic targets for treatment of rheumatoid arthritis. *Frontiers in Immunology, 12*, 686155. https://doi.org/10.3389/fimmu.2021.686155
8. Tariq, M., Ahmed, S., Al-Kuraishy, H. M., Al-Gareeb, A. I., & Elekhnawy, E. (2024). Identification of novel biomarkers in rheumatoid arthritis: A comprehensive review. *International Journal of Molecular Sciences, 25*(6), 2757. https://doi.org/10.3390/ijms25062757
9. Martens, M., et al. (2021). WikiPathways: Connecting communities via curated biological pathways. *Nucleic Acids Research, 49*(D1), D613–D621. https://doi.org/10.1093/nar/gkaa1024

10. Piñero, J., et al. (2017). DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research, 45*(D1), D833–D839. https://doi.org/10.1093/nar/gkw943
11. Ruiz-Noa, Y., et al. (2022). PADI4 haplotypes and rheumatoid arthritis risk: A meta-analysis. *Current Issues in Molecular Biology, 44*(9), 4252–4264. https://doi.org/10.3390/cimb44090292
12. Szklarczyk, D., et al. (2019). STRING v11: Protein–protein association networks with increased coverage. *Nucleic Acids Research, 47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131
13. Warde-Farley, D., et al. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research, 38*(Suppl. 2), W214–W220. https://doi.org/10.1093/nar/gkq337
14. Hu, H., Wang, J., Tang, J., Chen, X., & Deng, Z. (2022). A systems biology approach to identify the common signatures and potential influences of COVID-19 on rheumatoid arthritis. *Frontiers in Immunology, 13*, 860676. https://doi.org/10.3389/fimmu.2022.860676
15. Kuleshov, M. V., et al. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research, 44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377
16. Chen, E. Y., et al. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics, 14*(1), 128. https://doi.org/10.1186/1471-2105-14-128
17. Fabregat, A., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research, 46*(D1), D649–D655. https://doi.org/10.1093/nar/gkx1132
18. Mi, H., et al. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols, 8*(8), 1551–1566. https://doi.org/10.1038/nprot.2013.092
19. Shannon, P., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303
20. Huang, Y., et al. (2024). Identification of diagnostic genes in rheumatoid arthritis based on machine learning and immune infiltration analysis. *Frontiers in Immunology, 15*, 1431452. https://doi.org/10.3389/fimmu.2024.1431452
21. Antonia, R. J., Shirasaki, T., & Baldwin, A. S. (2022). STAT3 regulates inflammatory cytokine production in synovial fibroblasts through a feedforward loop. *Journal of Cellular and Molecular Medicine, 26*(16), 4591–4601. https://doi.org/10.1111/jcmm.17488
22. Wang, T., He, Q., & Chan, K. H. K. (2025). A multi-omics approach to identify and validate shared genetic architecture in rheumatoid arthritis, multiple sclerosis, and type 1 diabetes. *Molecular Genetics and Genomics, 290*. https://doi.org/10.1007/s10142-025-01598-x
23. Nair, A., et al. (2020). Chronic TNF exposure reprograms synovial fibroblasts to become inflammatory memory-like cells. *Cell & Bioscience, 10*(1), 1–16. https://doi.org/10.1186/s13578-020-00445-5
24. Anderson, A. E., et al. (2019). IL-6-driven STAT3-regulated genes in CD4+ T cells discriminate promising responders to treatment in early rheumatoid arthritis. *Rheumatology, 58*(7), 1250–1258. https://doi.org/10.1093/rheumatology/kez025
25. Gonzalez-Hormazabal, P., et al. (2024). Association of HLA-DRB1 alleles with rheumatoid arthritis susceptibility: A comprehensive meta-analysis. *Frontiers in Immunology, 15*, 1380921.

https://doi.org/10.3389/fimmu.2024.1380921

26. Burmester, G. R., & Pope, J. E. (2017). Novel treatment strategies in rheumatoid arthritis. *The Lancet, 389*(10086), 2338–2348. https://doi.org/10.1016/S0140-6736(17)31491-5
27. Zhang, F., et al. (2023). Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. *Nature, 623*, 616–624. https://doi.org/10.1038/s41586-023-06708-y
28. Lewis, M., Barnes, M., Blighe, K., et al. (2019). Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Reports, 28*(9), 2455–2470.e5. https://doi.org/10.1016/j.celrep.2019.07.091
29. Subramanian, A., et al. (2017). A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell, 171*(6), 1437–1452. https://doi.org/10.1016/j.cell.2017.10.049
30. The Gene Ontology Consortium. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research, 49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113
31. Clarke, D. J. B., Marino, G. B., Deng, E. Z., et al. (2024). Rummagene: Massive mining of gene sets from supporting materials of biomedical research publications. *Communications Biology, 7*, 482. https://doi.org/10.1038/s42003-024-06177-7
32. Ochoa, D., Hercules, A., Carmona, M., et al. (2023). The next-generation Open Targets Platform: Reimagined, redesigned, rebuilt. *Nucleic Acids Research, 51*(D1), D1353–D1359. https://doi.org/10.1093/nar/gkac1046
33. Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: Augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Research, 44*(D1), D380–D384. https://doi.org/10.1093/nar/gkv1277