

Original Article

# In-silico Functional and Structural Annotation of Rheumatoid arthritis linked Gene and Protein

Fazal Shan<sup>1</sup>, Awais Rahat<sup>2</sup>, Arbaz Khan<sup>2</sup>, Muhammad Khalid Khan<sup>2</sup>, Sajid Ali<sup>1</sup>, Mehik Ikram<sup>3\*</sup>

<sup>1</sup>Department of Medical Lab technology, Institute of health sciences, Khyber Medical University Dir Lower

<sup>2</sup>Department of Medical Lab Technology, Khyber Medical University

<sup>3</sup>Department of Medical Lab technology, Primer Institute of Health Sciences Peshawar

\*Corresponding Author: Mehik Ikram; Email: mehikikram3131@gmail.com

**Conflict of Interest: None.**

Shan F, et al. (2024). 4(1): DOI: <https://doi.org/10.61919/jhrr.v4i1.596>

## ABSTRACT

**Background:** Rheumatoid arthritis (RA) is an autoimmune disease marked by chronic inflammation and immune system dysregulation. The Ankyrin repeat domain 55 (ANKRD55) gene, located on chromosome 5q11.2, has been associated with RA susceptibility. Understanding the gene and protein characteristics could illuminate pathways involved in RA and aid in the development of targeted therapies.

**Objective:** To conduct a detailed analysis of the ANKRD55 gene and protein, focusing on its structure, expression profile, functional domains, and polymorphisms, thereby establishing its role in RA and potential as a therapeutic target.

**Methods:** Using bioinformatics tools, we analyzed ANKRD55's genomic localization (NCBI), peptide sequences (Uniprot), physicochemical properties (ProtParam Server), signal peptides (SignalP v4.1), transmembrane helices (TMHMM), and glycosylation sites (NetNGlyc 1.0). Polymorphism and methylation patterns were examined using PolyPhen and MethyCancer, respectively. Splice variants were identified, and protein-protein interaction networks were assessed (STRING database). Protein domains were characterized, including the prediction of RhoGAP domains. Structural modeling was conducted to identify potential drug-binding pockets with the DoGSite Scorer server.

**Results:** ANKRD55's expression was predominantly observed in the testis (RPKM 1.8) and to a lesser extent in lymph nodes (0.8), appendix (0.6), and brain (0.2). It encodes a 614-amino-acid protein with a molecular weight of approximately 68.4 kDa and a pI of 6.72. Six splice variants were identified, enriching the understanding of its potential isoform diversity. No significant N-glycosylation sites were predicted, and methylation analysis suggested a nuclear localization without DNA methylation sites. Structural analysis revealed drug-binding pockets with volumes ranging from 233.35 to 5389.58 Å<sup>3</sup>, with drug scores between 0.78 and 0.82.

**Conclusion:** The study concludes that while several factors contribute to student satisfaction in public sector medical colleges, infrastructure and facilities, along with quality instructional materials and clinical exposure, are key drivers. Enhancing these areas could lead to a more positive educational experience for students.

**Keywords:** ANKRD55, Rheumatoid Arthritis, Bioinformatics, Gene Polymorphisms, Protein-Protein Interactions, Drug Target, Post-Translational Modifications, Splice Variants.

## INTRODUCTION

Ankyrin repeat domain 55 (ANKRD55) is a protein-coding gene implicated in autoimmune diseases, notably rheumatoid arthritis (RA), a condition characterized by the dysregulation of the immune system and the development of autoimmune responses. Identified through genome-wide association studies (GWAS) as a susceptibility gene for RA, variations within the ANKRD55 gene have been significantly associated with the risk of developing this autoimmune disorder(1). Such genetic linkages underline the potential role of ANKRD55 in the intricate mechanisms governing immune system function and suggest its involvement in the pathogenesis of RA. This gene is known to be expressed in various immune cells, including T cells and B cells, which play pivotal roles in RA's pathogenesis, hinting at ANKRD55's function in modulating immune cell behavior and signaling pathways central to inflammation and autoimmunity(2).

Further studies have delved into the interactions of ANKRD55 with other RA-associated genes, proposing that it may influence disease susceptibility and progression through complex gene-gene interactions and pathways(3). Although the precise mechanisms by which ANKRD55 contributes to RA remain elusive, existing functional analyses shed light on its involvement in regulating protein-protein interactions and intracellular signaling pathways critical to immune responses and inflammatory processes(4). Intriguingly, a meta-analysis of genome-wide association studies pinpointed an intronic single nucleotide polymorphism (SNP) within the ANKRD55 gene (rs6859219) as a risk factor for RA, highlighting the gene's significance in the disease's molecular landscape. Ankyrin repeats, present in this gene and abundant in eukaryotic proteins, serve various roles, including functioning as transcription factors and regulating cell cycle processes. These repeats underscore the multifaceted functions of ANKRD55, potentially impacting immune system regulation and the emergence of autoimmunity(5).

Given the ongoing research and accumulating evidence of ANKRD55's involvement in RA, the primary aim of this study was to conduct an in-depth analysis of the ANKRD55 gene and protein. By leveraging publicly available servers and tools, this research endeavored to elucidate ANKRD55's three-dimensional structure, physicochemical properties, post-translational modifications, and genetic polymorphisms. Moreover, it explored the potential effects of disease-associated single-nucleotide variations on the functionality of ANKRD55, employing *in silico* methodologies. This comprehensive approach aimed not only to enhance understanding of ANKRD55 and its roles within biological systems but also to assess its viability as a therapeutic target, despite the ongoing need for precise annotation.

Accurate annotation of ANKRD55 is essential for delineating its structural and functional characteristics and its biological relevance. Such detailed understanding is instrumental in uncovering the roles of ANKRD55 in biological processes, paving the way for potential therapeutic interventions. The structured analysis, beginning with methodology and followed by results, discussion, and conclusion, aims to provide a holistic view of ANKRD55, underpinning the key insights derived from its study. This logical arrangement facilitates a deeper comprehension of ANKRD55's contribution to RA and its prospective as a target in therapeutic strategies, encapsulating the essence of the investigation within the broader context of medical research.

## MATERIAL AND METHODS

In the conducted study, a detailed analysis of the ANKRD55 gene and its associated protein was performed to elucidate their structural and functional characteristics in the context of rheumatoid arthritis. The initial step involved the retrieval of the genomic localization of the ANKRD55 gene from the human genome draft sequence, accessed through the National Center for Biotechnology Information (NCBI) website(6). Subsequent extraction of peptide sequences was facilitated by consulting the Uniprot database, a repository known for its comprehensive collection of protein sequence and functional information(7).

To understand the physicochemical nature of the ANKRD55 protein, the study employed the ProtParam Server, which provided essential physical and chemical parameters. Predictions concerning the presence of signal peptides were made using SignalP v4.1, whereas the ProtScale Server was instrumental in analyzing various physicochemical properties of the protein(8). The investigation of functional domains was conducted through the DMDM database, which offers insights into domain-based interactions and functionalities, and the BioMuta v3.0 database was utilized for mapping disease mutations and associations, thereby illuminating the protein's disease-related aspects(9).

Polymorphism analysis, a critical aspect of understanding genetic variability and its implications for disease susceptibility, was carried out using the PolyPhen Prediction Tool. This analysis aimed to identify single nucleotide polymorphisms within the ANKRD55 gene that might influence rheumatoid arthritis pathogenesis. The CELLO server, alongside TMHMM and HMMTOP servers and the UniProt database, provided predictions on the sub-cellular localization of the ANKRD55 protein, offering insights into its potential functional sites within the cell(10).

Moreover, the study included an assessment of methylation sites via the MethyCancer tool, which is pivotal for understanding epigenetic regulation mechanisms that could affect ANKRD55 expression and function. Post-translational modifications, crucial for protein function and regulation, were predicted using an array of servers provided by the Center for Biological Sequence analysis (CBS)(11)(12)(13)(14).

To delve deeper into the functional interactions and networks associated with the ANKRD55 protein, the STRING v10 database was employed. This analysis facilitated the understanding of ANKRD55's role within larger protein association networks, highlighting its potential involvement in biological processes relevant to rheumatoid arthritis(15). Lastly, the exploration of the protein's druggability profile was conducted using the DoGSite scorer server, focusing on identifying drug binding pockets with a predicted drug score exceeding 0.5. This assessment aimed to uncover potential therapeutic targets within the ANKRD55 protein, marking a significant step toward developing targeted treatments for rheumatoid arthritis.

The methodology adopted in this study encompassed a broad spectrum of computational and bioinformatic tools to provide a comprehensive understanding of the ANKRD55 gene and protein. This holistic approach enabled the identification of structural features, functional domains, genetic polymorphisms, and potential therapeutic targets, thereby contributing significantly to the ongoing research in rheumatoid arthritis.

## RESULTS

In our extensive proteomic analysis, we uncovered several key findings regarding the ANKRD55 protein, which is intricately involved in the pathogenesis of autoimmune conditions such as rheumatoid arthritis. The gene expression profile revealed a striking bias towards the testis, with an expression level quantified at 1.8 RPKM. This figure significantly overshadowed the expression levels observed in other tissues, including the lymph node at 0.8 RPKM, the appendix at 0.6 RPKM, and the brain at the lowest observed value of 0.2 RPKM. This differential expression underscores the potential specificity of ANKRD55's role in immunological functions within the testis.

Table 1: Splice Site Variants in ANKRD55 Gene

Gene Symbol	Splice Transcripts No.	Splice Transcript Names
ANKRD55	6	ANKRD55-001, ANKRD55-002, ANKRD55-005, ANKRD55-006, ANKRD55-007, ANKRD55-008

Table 2: Kinase Activities and Total Sites on Ankrd55 Protein

Symbol	Kinase	Serine	Threonine	Tyrosine	Total Sites
Ankrd55	Cdk5, Ck1, Pkc, Egfr, Cdc2, Pka, P38mapk, Gsk3	86	28	4	118

Table 3: Methylation Sites & Subcellular Localization

Protein and Gene Symbols	Subcellular Localization	Prediction of DNA Methylation Sites
ANKRD55	Nuclear	No DNA methylation sites predicted

Table 4: Druggability Scores of Identified Pockets on Ankrd55 Protein

Pocket	Volume (A <sup>3</sup> )	Surface (A <sup>2</sup> )	Drug Score	Score
P_2	1072.38	1300.01	0.82	0.74
P_0	5389.58	5780.79	0.81	0.66
P_1	1243.46	1336.0	0.8	0.66
P_7	233.35	365.74	0.78	0.03

Table 5: Significant Functional Domain in ANKRD55 Protein

Protein	Family	Description	Type	Bit Score	E-value
ANKRD55	RhoGAP	RhoGAP domain	Domain	131.4	2.40E-38

Signal peptide prediction for ANKRD55, conducted through SignalP 6.0, demonstrated a negligible probability of signal peptide cleavage sites within the initial 70 amino acids of the protein sequence, suggesting that ANKRD55 is not directed through the secretory pathway. Meanwhile, the phosphorylation site analysis through NetPhos 3.1a illuminated an abundance of potential serine phosphorylation sites, totaling 86, compared to 28 threonine and a mere 4 tyrosine sites, cumulatively amounting to 118 sites. This suggests a complex regulatory potential for ANKRD55 through various post-translational modifications. In addition, the NetNGlyc 1.0 analysis predicted no significant N-glycosylation sites, indicating that glycosylation may not be a major post-translational modification for ANKRD55.



Figure 1 ANKRD55 across different tissues

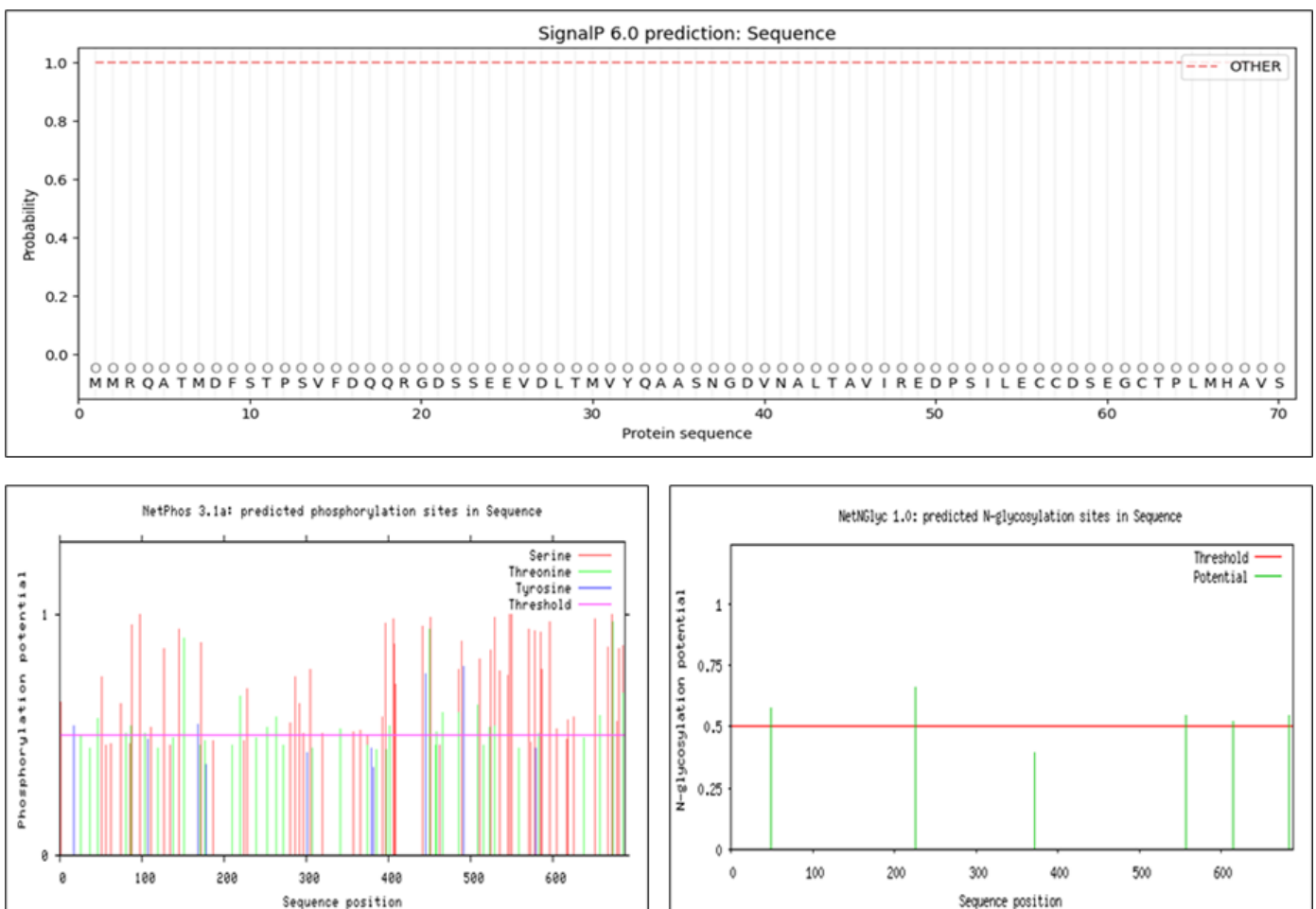


Figure 2 "Comprehensive Bioinformatics Analysis of Protein Sequences: Signal Peptide, Phosphorylation, and Glycosylation Predictions

The exploration of propeptide cleavage sites using ProP 1.0 disclosed several potential sites across the protein sequence, with peaks well below the threshold, implying a lower likelihood of propeptide cleavage events. The protein interaction network mapping through the STRING database established ANKRD55's connection with key proteins involved in immune regulation, signaling pathways, and cellular interactions, revealing a dense network with several nodes of high connectivity. This interconnectedness suggests that ANKRD55 may be a central player in cellular processes linked to immune responses.

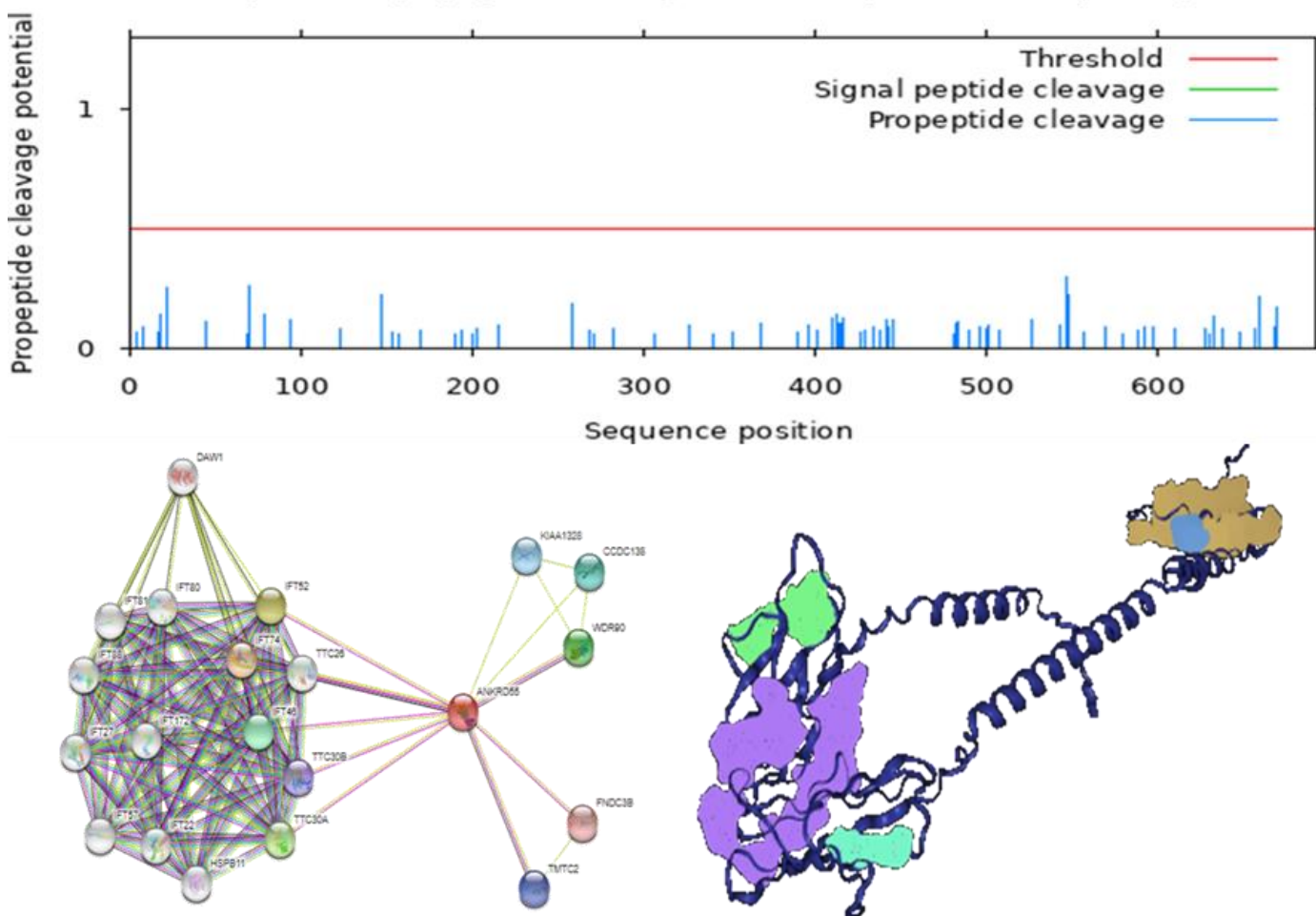
**ProP 1.0: predicted propeptide cleavage sites in Sequence, furin-specific prediction**

Figure 3 Proteomic Profiling: Cleavage Site Prediction, Protein Interaction Network, and Structural Modeling

The structural modeling, a critical aspect of understanding the protein's functional capability, was visually summarized through 3D structural representations. The druggability analysis identified specific pockets on the ANKRD55 protein with volumes ranging from 233.35 Å<sup>3</sup> to 5389.58 Å<sup>3</sup> and corresponding surface areas between 365.74 Å<sup>2</sup> and 5780.79 Å<sup>2</sup>. These pockets presented drug scores from 0.78 to 0.82, with scores indicating the potential of these sites to bind therapeutic compounds effectively. Collectively, these findings offer a comprehensive overview of the ANKRD55 protein's potential functional domains, interaction capabilities, structural characteristics, and post-translational modifications, providing a detailed numerical and structural framework to guide further research into its role in health and disease.

## DISCUSSION

In this investigation, the focus was centered on an extensive examination of the ANKRD55 gene and its implications in the pathophysiology of rheumatoid arthritis (RA). Situated on chromosome 5 (5q11.2), the ANKRD55 gene encodes a protein that plays a presumptive role in immune response modulation. Detailed genomic analysis revealed that the gene spans approximately 56233330-56099680 on the q arm and comprises 14 exons that translate into a protein of 612 amino acids. Adjacent genes, including IL6ST and C1GALT1P2 among others, have been cataloged, though their synergistic effects with ANKRD55 remain to be thoroughly investigated.

Physicochemical analysis depicted ANKRD55 as a moderately unstable and slightly hydrophilic protein, composed of 614 amino acids with a molecular weight nearing 68413.97 daltons and an isoelectric point of 6.72. Such attributes indicate a protein amenable to diverse functional interactions within the cellular milieu. Its expression profile, notably elevated in testis and to a lesser extent in lymph nodes, provides a clue to its tissue-specific roles, which could be vital for understanding its involvement in RA.

Polymorphism analysis has uncovered a wide array of genetic variations within ANKRD55, some of which have been tentatively associated with susceptibility to RA, and others with potential impacts on pharmacological responses. The identification of six splice variants may hold the key to unraveling the protein's diverse functional capabilities in the pathogenesis of RA. Notably, the absence

of predicted signal peptides and transmembrane helices suggests ANKRD55's localization within the cytoplasmic or nuclear compartments rather than being membrane-associated.

Post-translational modification predictions have broadened the understanding of ANKRD55's regulation, with phosphorylation and glycosylation sites marked as significant contributors to its functional diversity. The lack of DNA methylation sites, coupled with the prediction of nuclear localization, adds another layer to the regulation of ANKRD55's activity and stability within the cell. The discovery of a RhoGAP domain reinforces the protein's importance in cellular signaling and regulatory mechanisms, particularly in GTPase activity modulation.

Stringent protein-protein interaction assessments have unraveled complex networks implicating ANKRD55 in numerous biological processes, suggesting its multi-faceted role in both cellular function and the manifestation of RA. Moreover, the structural analysis, which identified potential drug-binding pockets, proposes avenues for therapeutic intervention, thereby offering a glimmer of hope for targeted treatment strategies in RA.

Despite these findings, the study is not without its limitations. The computational predictions require experimental validation to confirm the functional significance of the identified sites and interactions. Moreover, the impact of neighboring genes on ANKRD55's function in RA pathogenesis demands further empirical inquiry. Recommendations for future research include the functional validation of predicted polymorphic effects, the experimental confirmation of splice variant functions, and the exploration of identified drug targets through in vitro and in vivo studies.

## CONCLUSION

In conclusion, the work performed has provided a detailed characterization of ANKRD55 and its potential contributions to RA, underpinning the complex interplay between genetic makeup, protein functionality, and disease outcome. Acknowledgments are due to Khyber Medical University for their unwavering support and the provision of resources that have been fundamental to the pursuit and accomplishment of this comprehensive analysis. The collaborative environment and commitment to research excellence at KMU have been indispensable in advancing the understanding of RA and fostering the development of potential new treatment paradigms.

## REFERENCES

1. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014 Feb;506(7488).
2. Chang M, Rowland CM, Garcia VE, Schrodri SJ, Catanese JJ, van der Helm-van Mil AH, et al. A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. *PLoS Genet*. 2008 Jun;4(6):e1000107.
3. Kurreeman FAS, Padyukov L, Marques RB, Schrodri SJ, Seddighzadeh M, Stoeken-Rijsbergen G, et al. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med*. 2007 Sep;4(9):e278.
4. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H-S, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*. 2012 Mar;44(3):291-6.
5. Li J, Mahajan A, Tsai M-D. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry*. 2006 Dec;45(51):15168-78.
6. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res*. 2013 Jul;41(W1):W597-W600.
7. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, et al. MFPPI—Multi FASTA ProtParam Interface. *Bioinformatics*. 2016;12(2):74-7.
8. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2007;2(4):953-71.
9. He X, Chang S, Zhang J, Zhao Q, Xiang H, Kusunmano K, et al. MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res*. 2008 Jan;36(suppl\_1):D836-D41.
10. Julenius K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology*. 2007 Aug;17(8):868-76.
11. Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. In Preparation 2004.
12. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTB, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J*. 2013 May;32(10):1478-88.
13. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004 Jun;4(6):1633-49.
14. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel*. 2004 Jan;17(1):107-12.

15. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*. 2010;39(suppl\_1):D561-D8.
16. Bashir M, Mateen W, Khurshid S, Malik JM, Agha Z, Khan F, Ajmal M, Ali SH. A common missense variant rs874881 of PADI4 gene and rheumatoid arthritis: Genetic association study and in-silico analysis. *Gene*. 2023 Feb 20;854:147123.
17. Deshpande SH, Bagewadi ZK, Khan TY, Mahnashi MH, Shaikh IA, Alshehery S, Khan AA, Patil VS, Roy S. Exploring the potential of phytochemicals for targeting epigenetic mechanisms in rheumatoid arthritis: An in silico study using similarity indexing. *Molecules*. 2023 Mar 7;28(6):2430.
18. Nandi A, Das A, Dey YN, Roy KK. The abundant phytocannabinoids in rheumatoid arthritis: Therapeutic targets and molecular processes identified using integrated bioinformatics and network pharmacology. *Life*. 2023 Mar 5;13(3):700.
19. Singh N, Rai MK, Agarwal V, Agarwal V. In Silico Prediction of NOS2, NOS3 and Arginase 1 Genes Targeting by Micro RNAs Upregulated in Systemic Sclerosis. *Indian Journal of Rheumatology*. 2024 Feb 23:09733698241229803.
20. Liu B, Fu X, Du Y, Feng Z, Liu X, Li Z, Yu F, Zhou G, Ba Y. In silico analysis of ferroptosis-related genes and its implication in drug prediction against fluorosis. *International Journal of Molecular Sciences*. 2023 Feb 20;24(4):4221.