Readers Insight
Publisher of open access books and journals
PJMLS
Print ISSN: 2707-4471. Online ISSN: 2707-448X

| Research Article | Pak-Euro Journal of Medical and Life Sciences |
|---|---|
| DOI: 10.31580/pjmls.v6i4.3011 | Copyright © All rights are reserved by Corresponding Author |
| Vol. 6 No. 4, 2023: pp. 527-540 | |
| www.readersinsight.net/pjmls | **Revised:** December 13, 2023 **Accepted:** December 25, 2023 |
| **Submission:** October 07, 2023 | **Published Online:** December 31, 2023 |


# IN-SILICO STRUCTURAL AND FUNCTIONAL ANNOTATION OF FCGR2A IN ASSOCIATION WITH RHEUMATOID ARTHRITIS AND IDENTIFICATION OF NOVEL THERAPEUTIC TARGETS

Fazal Shan[1]*, Ahmad Al Abad[2], Majid Ullah[3], Nighat Nawaz[4], Mihk Ikram[1], Ishtiaq Ahmad[5], Irshad Ahmad[1]*

[1]**Department of Molecular Biology and Genetics,** Khyber Medical University Peshawar, Pakistan

[2]**Department of Pathology,** Bannu Medical College, Bannu, Pakistan

[3]**Department of Medical Lab Technology,** Khyber Medical University Peshawar, Pakistan

[4]**Department of Chemistry,** Islamia College, Peshawar, Pakistan

[5]**Internal Medicine Resident,** Geisinger Wyoming Valley Medical Center Wilkes-Barre, PA, United States.

***Corresponding Authors:** Fazalshan. E. mail. fazalshan.ibms@kmu.edu.pk and Dr. Irshad Ahmad. E. mail. irshadahmad@kmu.edu.pk

**Abstract**

*Rheumatoid arthritis (RA) is a chronic autoimmune disorder characterized by inflammation and joint destruction. Various genetic studies have identified 100 susceptible genes linked with RA. Among them, four SNPs (rs6671847, rs1801274, rs17400517, and rs6668534) in the FCGR2A gene have been shown to be associated with RA susceptibility. The functional consequences of these pathogenic variants on the protein structure, stability, and interaction with other molecules remain unknown. In this study, we used multiple in silico methods to examine the effects of these mutations on the FCGR2A gene. We annotated the gene structurally, and functionally, and Predicted physiochemical characterization, gene expression profiling, and Post-translational modification (PTM) analysis. We found that the (rs1801274 H167R) mutation in FCGR2A had a significant effect on the protein structure and interactions, whereas the other three SNPs (rs6671847, rs17400517, and rs6668534) had no significant impact on protein function. We also performed protein-protein interaction and molecular docking studies to identify potential therapeutic targets for RA. Our structure-based analysis showed that these mutations affect the protein core regions and functional domains, leading to altered protein-protein interactions both directly and indirectly. We identified three novel potential target sites for the development of drugs to treat RA by developing a drugability profile. These findings may have significant implications for the pathogenesis of RA and can be validated further through in vitro studies. Our study demonstrates the potential utility of multidirectional computational analysis for screening clinically relevant mutations in RA before undertaking biological assays. The findings highlight the importance of understanding the structural and functional consequences of pathogenic variants for developing effective therapeutic strategies. Overall, the insights gained from our in-silico approaches could aid in the development of novel therapies for RA and could pave the way for modified medicine in the treatment of this complex disease.*

**Keywords:** *Druggability profile, Gene expression, In silico methods, Molecular docking, Physiochemical characterization, Protein structure, Post-translational modification, Protein-protein interaction, Rheumatoid arthritis, Single nucleotide polymorphism.*

## INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic, systemic autoimmune disorder with a global prevalence of approximately 1%. The disorder is associated with an increased risk of disability due to joint inflammation and subsequent articular damage, which may result in permanent joint dysfunction (1). The exact cause of

RA is still not fully understood, but Some risk factors including environment, gender, and oral contraceptive are considered the cause of RA. Studies suggest that both genetic and environmental factors play a significant role in the development of RA (2, 3). Genetic factors are estimated to contribute around 60-65%, which highlights the significant contribution of genetic factors in disease development (4). Genome-wide association studies have identified approximately 100 gene locations that are associated with the disease. The most significant genetic association with RA is with the Human Leukocyte Antigen (HLA), which accounts for approximately 30% of the genetic contribution. Additionally, non-HLA genes have also been found to be significantly associated with the development of RA (5). Some other studies in different populations identify numerous susceptible genes that collectively explain around 15% of the disease variance (6-8). Among the RA-susceptible genes, the FCGR2A gene has been recognized as a major contributor to RA. FCGR2A is a member of the FC gamma receptor (FCGR) family, which plays a crucial role in the recognition of immune complexes (ICs) by the immune system (9, 10). The advent of genome-wide association studies has significantly improved our understanding of the genetic basis of RA by identifying variants associated with the disease. Recent study has highlighted the significant role of specific genetic variants in the FCGR2A gene (rs6671847, rs1801274, rs17400517, and rs6668534) in increasing the risk of developing RA (11). This study aimed to use computational approaches to predict the effects of missense variants on the amino acid sequence and conservation and the structural and functional properties of proteins implicated in rheumatoid arthritis. By analyzing FCGR2A SNPs, best potential therapeutic sites might be predicted for the development of new drugs to treat RA. In-silico functional and structural annotation of FCGR2A variants is a crucial step in the development of novel therapeutic strategies for RA.

# MATERIALS AND METHODS

## SEQUENCE RETRIEVAL

The localization of genes associated with rheumatoid arthritis (RA) on the human genome was determined by analyzing the draft sequence of the human genome available on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/) (12). The peptide sequences of human proteins associated with rheumatoid arthritis (RA) were obtained from the UniProt (http:// www.unipro ot.org/) database, which is a comprehensive resource of protein sequences and functional information.

## SEQUENCE BASED PHYSICO CHEMICAL CHARACTERIZATION

The Protparam Server (http://web.expasy.org/protparam/) tool available on the ExPASy website, was utilized to compute the physical and chemical properties of the human peptide sequence associated with RA. This includes the calculation of various parameters such as the molecular weight, isoelectric point, and amino acid composition. The server was also used to analyze the structural and functional motifs of the peptide sequence, which may provide information about its potential biological activity and interactions with other molecules (13).

## FUNCTIONAL DOMAIN ASSESSMENT AND POLYMORPHISM ANALYSIS

The Domain Mapping of Disease Mutations (DMDM) database (http://www.bioin f.umbc.edu/dmdm/), was utilized to investigate the functional domains of proteins associated with RA. The DMDM database contains information on disease-related mutations that may affect the structure and function of protein domains (11). The database maps genetic polymorphisms in the protein domains. BioMuta v3.0 database (https://hive.biochemist ry.gwu.edu/bio muta) was used to check the disease associated with genetic polymorphism.

## PREDICTION OF POST TRANSLATIONAL MODIFICATIONS (PTMS)

The Center for Biological Sequence Analysis (CBS) (http://www.cbs.dtu.dk/services/) provides a variety of tools and servers for the prediction and analysis of post-translational modifications (PTMs) on proteins. The available servers NetCGlyc 1.0(14), NetCorona 1.0(15), NetGlycate 1.0 (16), NetNGlyc 1.0 (17), NetOGlyc

4.0(18), NetPhos 3.1(19), and ProP 1.0(20) servers were utilized to predict potential sites of glycosylation, mannosylation, and phosphorylation on the peptide sequence of the protein associated with RA.

## SELECTION OF RA SUSCEPTIBILITY LOCI

In this study, different types of genetic variation, including variants, single nucleotide polymorphisms (SNPs), and mutations, were analyzed to confirm the RA susceptibility loci of FCRGA2. The genetic variation table provided by the Ensemble project was used to obtain relevant information such as rsID, genome-wide significance, and odd ratios of the FCRGA2 gene in relation to RA. The analysis specifically focused on missense mutations to explore the potential impact of amino acid substitutions on the function of the FCRGA2 protein in RA

## IN SILICO ASSESSMENT OF HARMFUL POTENTIAL OF NSSNPS OF RA

In this study, a variant effect predictor (VEP) toolset was utilized to analyze and annotate genomic variants in the coding regions of the genome. VEP is a powerful computational tool that provides easy access to various genetic annotations and supports different input data formats such as the VCF, dbSNP, and HGVS formats. The results of the analysis provide detailed information about the chromosomal location, variant effect, transcript ID, and predicted functional impact of each variant, including the scores generated by the SIFT, PolyPhen-2, CADD, and FATHMM algorithms.

## STRUCTURAL ANALYSIS OF RA PROTEINS

### *3-DIMENSIONAL MODELING OF WILD TYPE AND MUTANT PROTEIN MODELS*

In this study, the crystal structure of a normal protein was obtained from the Protein Data Bank (PDB) (www.rcsb.org/) and used as a template to generate 3D structures of a mutant sequence using the Swiss Model online protein prediction server (https://swissmodel.expasy.org/). (21) The best model was selected based on various scores such as the estimated template model score, confidence score, and root mean square deviation (RMSD) scores. The quality of the generated structure was further analyzed by building a Ramachandran plot using the Rampage server. (22)

### *SECONDARY STRUCTURE ANALYSIS*

To found an association between mutant amino acid sequences and protein structures, information about variations in different secondary structural elements such as the location of helices (a), strands (b) bends, and turns. This information can be obtained from the PDB-Sum database, which provides information on the fundamental structural elements of protein scaffolds. By analyzing this information, it is possible to gain insight into how mutations may affect protein structure and function.

### *PROTEIN STRUCTURAL SUPERIMPOSITION*

The two complementary methods such as mCSM (mutation Cutoff Scanning Matrix) and SDM (Site Directed Mutator) into a consensus/optimized prediction, using support vector machines (SVMs) was used to comprehend the effect of point mutations on the protein structure for this purpose we used DUET, bioinformatics server. Wild protein structure from PDB was used as input and additionally to amino acid information (wild type and mutant) in one letter codes.

## FUNCTIONAL PROTEIN ASSOCIATION NETWORKS

The functionality of a biomolecule is influenced by its interactions with different partners. The Search Tool for Retrieval of Interacting Genes and Proteins (STRING v 10) is a database that provides access to cluster of orthologs analysis servers and protein-protein interaction resources.(23). In this study, the database was used to identify the associations and interactions of RA proteins with other proteins and biomolecules.

## PROTEIN-PROTEIN MOLECULAR DOCKING

This study utilized Cluspro docking server to investigate the interaction potential and structural plasticity between RA proteins and their interacting partners. The ligand molecule was selected based on the highest confidence score of the queried protein in the molecular network obtained from the STRING database, while the query protein was considered as receptor. The molecular model of protein obtained from the PDB or Swiss model for mutant structure were used, and the default setting of the "free energy calculation" was set at 180_ to enable the maximum possible rotation for receptor and ligand sampling of their individual centroids. Several parameters were set, including a network spacing of 0.6 Å, positive and negative steps of 0.75 Å, 53 intermolecular partings intended in 20 steric scan stages. The final 25 phase was applied to obtain the highest orientation score of 0.76.2 Å, and 500 clusters were retained from the best 1000 orientations to obtain 10,000 the lowest ordered docking energy score.

## PREDICATION OF DRUGABILITY PROFILE

The druggability of selected proteins was evaluated by predicting the number and quality of their drug-binding pockets. The DoGSite scorer server was employed for this purpose, which provided information about the size, volume, hydrophobicity and shape of the predicted pockets, along with corresponding drug scores. The pockets with drug scores higher than 0.5 were considered to have high drugability potential and were selected for further analysis. In cases where multiple pockets met the cutoff score, the top three pockets were chosen.

# RESULTS

## SEQUENCE RETRIEVALS OF FCGR2A

The FCGR2A gene has a length of 18618 nucleotides and consists of 12 exons, which ultimately encode 317 amino acids. The neighboring genes include L0C105371473, L0C115801453, TRD-GTC10-1, TRG-GCC2-1, HSPA6, and TRL-CAG1-6. Located on the p arm of human chromosome 1 (1q23.3), FCGR2A occupies the genomic position between 161505430 and 161524048. A representation of the chromosomal location, neighboring genes, corresponding position, and exon structure of FCGR2A is presented in Fig. 1 (a, b, c).

## SEQUENCE-BASED PHYSIO-CHEMICAL CHARACTERIZATION OF FCGR2A

FCGR2A was subjected to predict sequence-based physio-chemical characterization by using protparam, which revealed that it has a molecular weight of 35000.68 with speculated formula $C^{1546}H^{2427}N^{425}O^{474}S^{14}$, total number of atom is 4886 and an isoelectric point of 6.19. Its aliphatic index was determined to be 81.51, and its instability index indicates that it is an unstable protein. The grand average of hydropathicity (GRAVY) value was -0.307. Its half-life in mammalian reticulocytes was estimated to be >30 hours in vitro, >20 hours in yeast, > 10 hours in Escherichia coli in vivo.

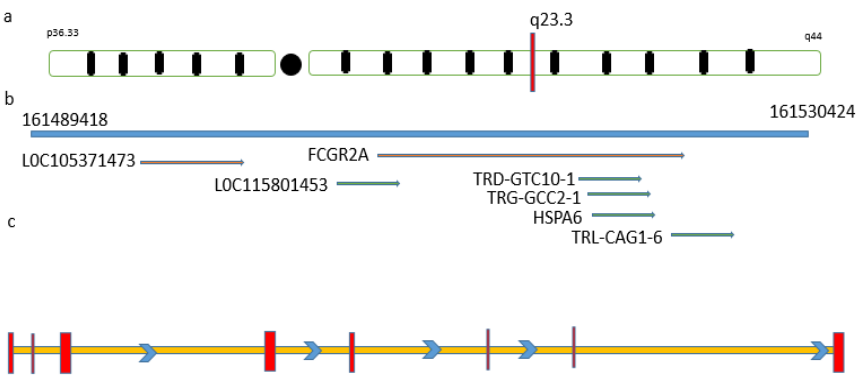



**Fig. 1.** FCG2A gene location and structure

## DOMAIN ASSESSMENT AND POLYMORPHISM

DMDM discovered one Single nucleotide providing a functional domain of FCGR2A. Four main domains were discovered in the protein (IG_like, IGc2, IG, and IG_like). IG_like domain on 46-119, 127-205 Immunoglobulin C type 2 (IGc2) domain 53-110 both are responsible for binding of FC portion of the IgG antibody and contribute in downstream signaling in cell. Immunoglobulin (IG) Domain 127-205 (Fig. 2 and Table I).

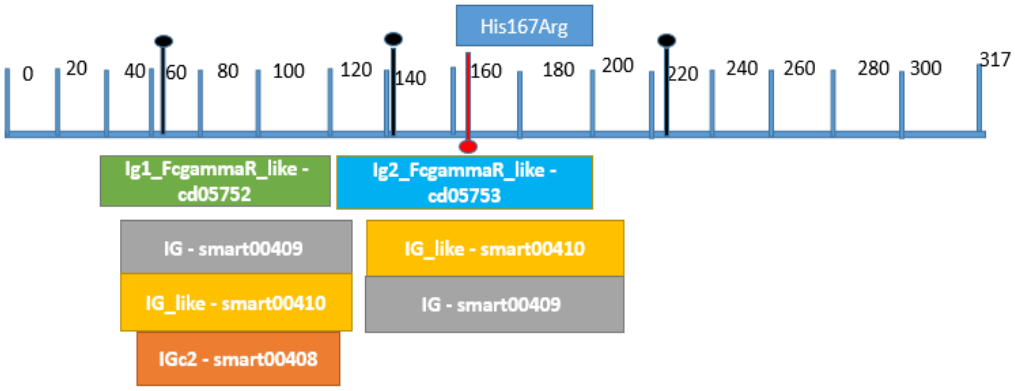



**Fig. 2.** FCGR2A domain and SNPs assessment

## POST TRANSLATIONAL MODIFICATIONS OF FCGR2A

The CBS server for Biological Sequence Analysis was employed to forecast various post-translational modifications on the proteins associated with RA. The default cutoff value of 0.5 was used, with values greater than 0.5 indicating a higher likelihood. The cleavage site of the propeptide signal (arginine/lysine) was estimated using the ProP 1.0 prediction server, which was found to be between the 32nd and 33rd amino acid Fig. 3a. The NetCGlyc 1.0 server was utilized to calculate the C-mannosylation sites, and it revealed one C-mannosylation site in the protein. In addition, NetGlycate 1.0 server predicted 13 glycation sites, NetNGlyc1.0 server predicted 2 N-linked glycosylation sites (Fig. 3b) and NetOGlyc 4.0 server predicted 1 serine-containing O-linked glycosylation site (Fig. 3c). The NetPhos 3.1 server predicted a total of 39 phosphorylation sites, with serine residues hosting the most (20), followed by threonine (14) and tyrosine (5) sites (Fig. 3d). The sites and their respective kinases are Sites with their respective kinases are CDK5, PKC, DNAPK, P38MAPK, GSK3, SRC, INSR, PKA, and EGFR. While Figures 3.4a-d provide graphical representations of the cleavage glycosylation, and phosphorylation sites predicted.

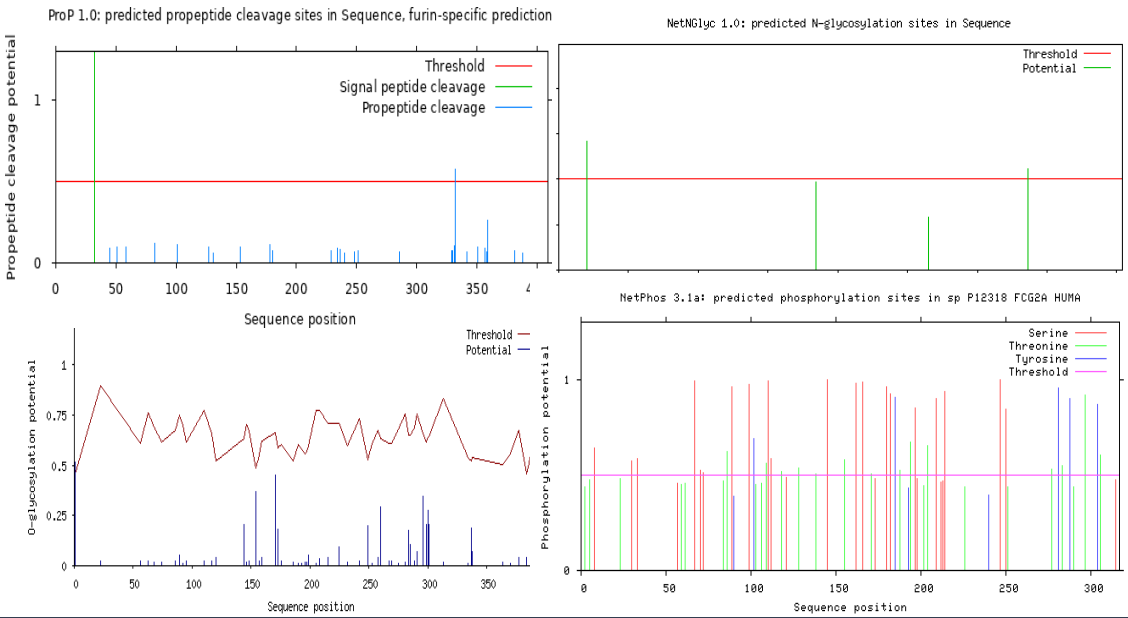


**Fig. 3.** Post-translational modifications on the FCRGA2 protein. The default cutoff value of 0.5 was used, with values greater than 0.5 indicating a higher probability. **A** ProP 1.0 prediction server cleavage site. **B.** NetOGlyc server predicted 2 N-glycosylation sites. **C.** NetOGlyc 4.0 server predicted 1 serine-containing O-linked glycosylation site. **D.** NetPhos 3.1 server predicted a total of 39 phosphorylation sites

## CODING SNPS OF RA RISK LOCI

Current study present molecular information about coding single nucleotide polymorphisms (SNPs) associated with the risk of RA, including the name of the affected gene, its chromosomal position, the change in the complementary DNA (cDNA), and the type and position of the amino acid variant resulting from the SNP (Table II ).

**Table I.** FCGR2A domain assessment

| Protein | Domain | Domain Accession | Start | END | E-value |
|---|---|---|---|---|---|
| FCRGA2 | IG_like | smart00410 | 46 | 119 | 0.0002 |
| | IGc2 | smart00408 | 53 | 110 | 9.9e-05 |
| | IG | smart00409 | 127 | 205 | 6.5e-10 |
| | IG_like | smart00410 | 127 | 205 | 6.5e-10 |

**Table II.** Coding SNPs of RA risk loci

| Gene | Variant id | Genomic location | Alleles | Consequence type | Resulting amino acid | cDNA location | Protein Effect |
|---|---|---|---|---|---|---|---|
| FCGR2A | rs6671847 | 1:161509020 | G/A | Intron variant | - | - | - |
| | rs1801274 | 1:161509955 | A/C/G | missense variant | H/R | c.500A>G | H167R |
| | rs17400517 | 1:161515469 | G/T | Intron variant | - | - | - |

## DELETERIOUS POTENTIAL OF MISSENSE OF MUTATIONS

For pathogenic effect on amino acid substitutions based on their deleterious potential on the function of candidate protein the SIFT prediction algorithm was used and basically it work on nucleotide homology principle. The values range from 0 to 1, lower score indicate high deleterious effect of unknown mutation towards the structural and functional features of corresponding protein. Missense mutation HIS167Arg, (SIFT score is 0) discovered it's tolerated effect on the function of FCGR2A protein while others SNPs were unpredicted. For the pathogenicity effect on amino acid substitution mutation, Polyphen algorithm was used. Naive Bayes probabilistic score to calculate the pathogenicity potential. The HIS167Arg, (score is 0.001) missense mutation of the FCGR2A gene presented a likely benign effect whereas are also found unpredicted. To classify the variant on the basis of non-functional variants (c-score is 10%), damaging variants (c-score is 20%), and lethal (c-score is 30%), all the required information was obtained from CADD. which is an integrative annotation of multiple mutation prediction methods into one framework. As per score percentage variant His167Arg are damaging variants which very near to the cuff of value (c-score is 20%).

To check the functional effects of coding variants in the form Rare Exome Variant ensemble learner REVEL tool will be used to determine the impact caused by amino acid substitution on the biological function of the protein. The cut off value of prediction is 0.5. If the value is more than 0.5 it showed likely disease causing effect and value is less than 0.5 it showed Likely benign of the variant. According above standard REVEL also predicated the Likely benign effect of protein (Table III).

**Table III.** Deleterious potential of missense

| Variant | Pathogenicity predictions | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SIFT | | PolyPhen | | CADD | | REVEL | |
| | Score | Predictions | Score | Predictions | Score | Predictions | Score | Predictions |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| rs6671847 | - | - | - | - | - | - | - | - |
| rs1801274 | 0.21 | Tolerated | 0.001 | Benign | 0 | Likely benign | 0.053 | Likely benign |
| rs17400517 | - | - | - | - | - | - | - | - |

## PROTEIN STRUCTURAL ANALYSIS

### *FCGR2A AND NFKBIE MUTANT PROTEIN MODELING AND STRUCTURAL*

### *FCGR2A 3D MODELING*

The crystal structures of the wild-type FCGR2A protein were obtained from the PDB database. A mutant form was created by introducing amino acid changes into the protein sequence, and a model structure of the FCGR2A protein was developed through Swiss Model. Among the various predicted structures, the best structures were selected based on their high Global Model Quality Estimation (GMQE) scores, which range from 0 to 1 and indicate reliability (Fig. 4). Homologous structures of these mutant proteins were selected for validation through Ramachandran plot assessments to evaluate their stereochemical quality (Fig. 5).

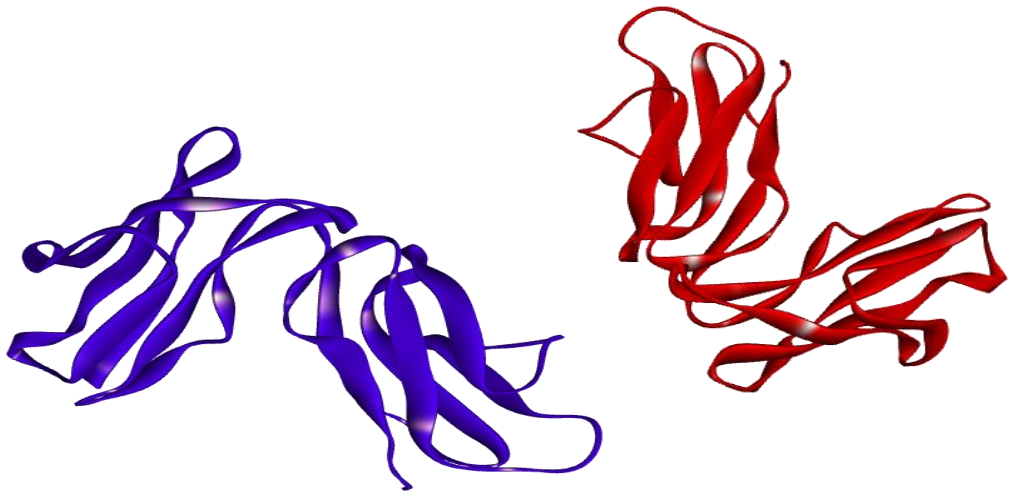
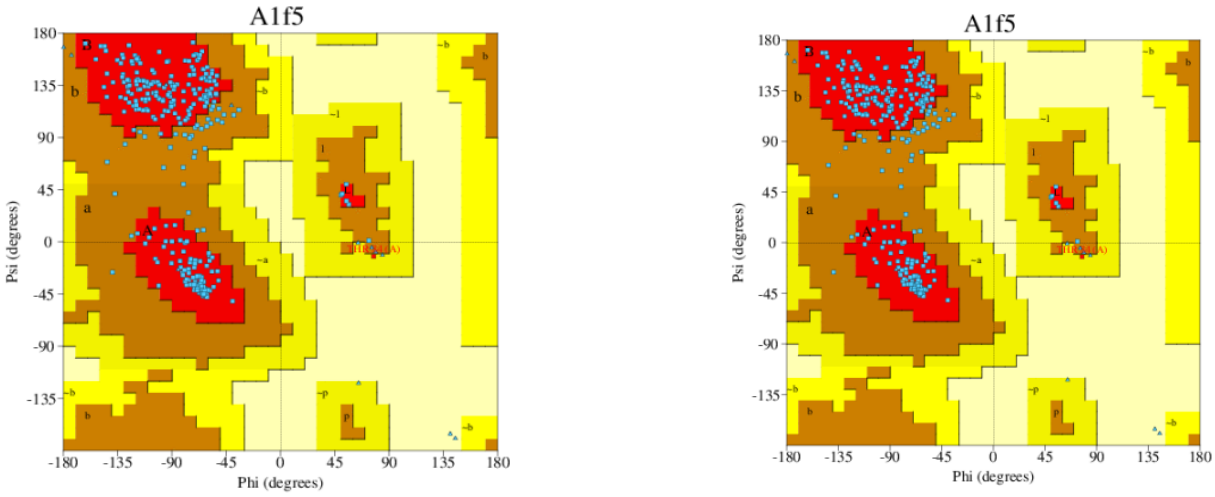


**Fig. 4.** 3D-Structural image of the developed protein models of (1) wild (Red in color) H167R (blue in color) generated from SWISS-MODEL

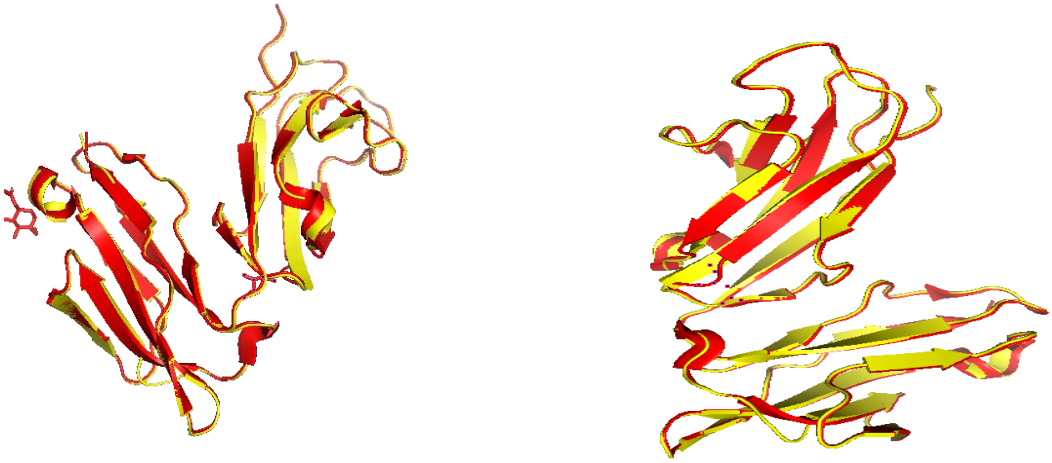



FCGR2A H167R

**Fig. 5.** FCGR2A and H167R Ramachandran plot generated for the wild and mutant protein of FCGR2A constructed structure. Vast majority of the amino acids satisfy the set constrains for the $\Psi$ and $\Phi$ angles

## *FCGR2A SECONDARY STRUCTURE AND SURROUNDING AMINO ACID CHANGES ANALYSIS*

In order to investigate the impact of amino acid substitutions on secondary structural features, we analyzed the structural differences in elements such as sheets (b-sheets, b-hairpins, b-bulges, and strands),



helices (6 helix–helix interactions), and turns (b turns, g turns, and disulfide bonds) between the wild and mutant models of FCGR2A. The substitution of the residue at position 167 in FCGR2A (H167R) resulted in 4 sheets, 4 b-hairpins, 3 b-bulges, 7 helices, 6 helix–helix interactions, 12 g-turns, 17 strands, 26 beta turns, 12 gamma turns, and 2 disulfides, with all elements in the secondary structure of the protein remaining similar to the wild type. The results are depicted in Fig. 6.

**Fig. 6.** Secondary structure analysis

**FCGR2A wild** **FCGR2A mutant**

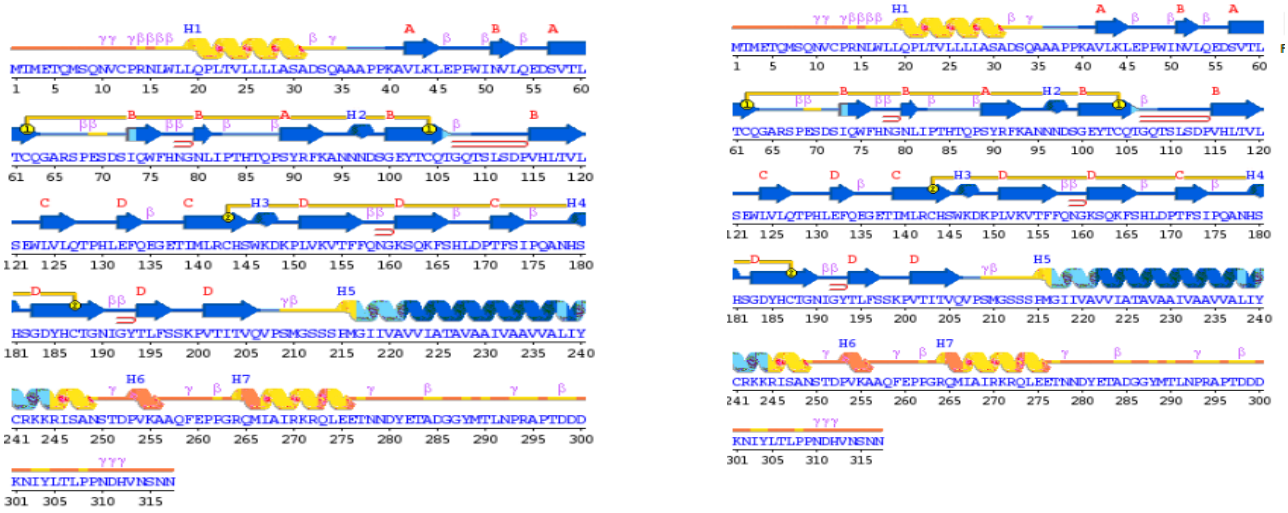


**Fig. 7.** Overall some changes was seen during the superimposing of wild and mutant structures of FCGR2A(red color is normal and yellow is mutant structure)

## *PROTEIN STRUCTURAL SUPERIMPOSITION*

The two complementary methods such as mCSM (mutation Cutoff Scanning Matrix) and SDM (Site Directed Mutator) into a consensus/optimized prediction, using support vector machines (SVMs) was used to comprehend the effect of point mutations on the protein structure for this purpose we used DUET, bioinformatics server. Wild protein structure from PDB was used as input and additionally to amino acid information (wild type and mutant) in one letter codes (Fig. 7).

## *PROTEIN – PROTEIN INTERACTION (PPI)*

The study explored the protein association network of FCGR2A protein using the STRING database to identify its physical interacting partners in the cellular context related to rheumatoid arthritis (RA). Analysis of the network revealed a limited association of FCGR2A with ten proteins including SYK, ITGAM,

CRP, LYN, HCK, FYN, ITGB2, SRC, FGR, and LAT, with the highest interaction score (c-score=0.998) observed for SYK. The predicted interaction networks suggest that FCGR2A and its protein partners activate various autoimmune reactions that are central to the pathogenesis of RA (Fig. 8).

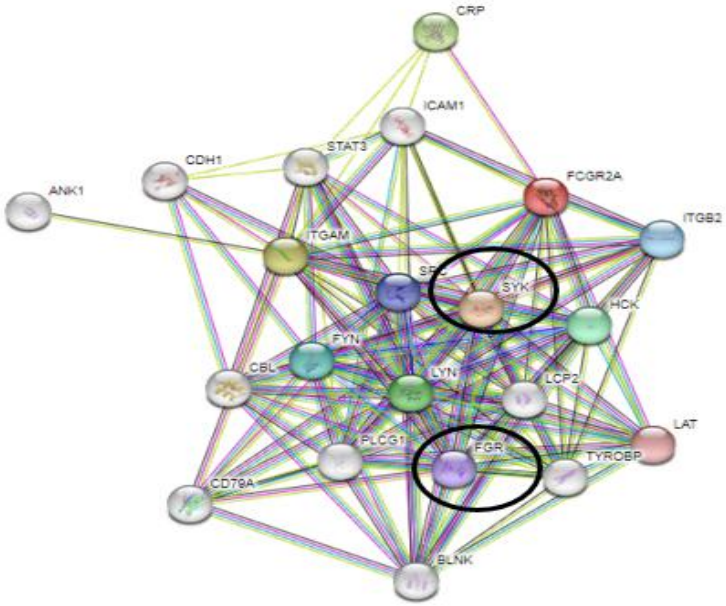



**Fig. 8.** FCGR2A Protein interaction network of FCGR2A STRING web server. Black circles indicate query protein and best interacting partner protein

## MOLECULAR DOCKING OF FCGR2A

The protein networking analysis revealed the key interacting partners of FCGR2A and their molecular functions, which were further investigated for the impact of missense mutations on these interactions. The mutations led to changes in the interacting amino acid residues, resulting in altered interactions with ligand molecules. Among the proteins analyzed, FCGR2A showed the highest binding energy (-179.7+/-13.1 Kcal/ Mol) with the ligand molecule SYK, whereas FCGR2A (H167R) exhibited the lowest binding energy with the same ligand molecule (-155.0+/-24Kcal/ Mol). In the wild-type FCGR2A, hydrophobic interactions with ligand molecule HDAC1 were formed by amino acids (table 3.10). However, in the mutant FCGR2A (H167R), different amino acids, participated in the formation of hydrogen bonds with the ligand molecule SYK. These findings suggest that the missense mutations may affect the interactions of FCGR2A with ligand molecules and thereby impact its molecular function (Fig. 9 & Table IV).

Table IV. The molecular docking scores

| **Protein complex** | | **Binding energy (Kcal/Mol)** | **Energy difference** | **Interacting amino acid** | |
|---|---|---|---|---|---|
| **Protein** | **Ligand** | | | **Receptor FCGR2A** | **Ligand SYK** |
| WILD | | -179.7 | | LYS166 | THR35 |
| FCGR2A | SKY | | | SER149 | GLN112 |
| | | | 014.7 | VAL172 | THR111 |
| | | | | THR171 | GLY116 |
| | | | | ARG3 | ASP80 |
| | | | | LYS166 | THR103 |
| | | | | LYS160 | ASN36 |
| | | | | VAL168 | TYR152 |
| His 167 Arg | | -155.0 | | Glu132 | LYS194 |
| | | | | ASN80 | GLN107 |
| | | | | THR118 | GLU89 |

| | |
|---|---|
| LYS199 | ARG4 |
| LYS199 | ALA5 |
| THR202 | GLN86 |
| TRP15 | MET113 |
| GLU132 | HIS90 |
| THR202 | ASP82 |
| ASN70 | TRP15 |


**Fig. 9.** Wild FCGR2A and mutant FCGR2A docking with SYK

## STRUCTURE BASED DRUGABILITY PROFILE ASSESSMENT WILD AND MUTANT FCGR2A

Protein structure is critical for understanding protein function, and advanced tools such as the protein pocket discovery tool DoGSite scorer (protein-plus) can aid in this endeavor. We utilized this tool to detect pockets with drug scores greater than 0.50, and then selected the best pockets based on a generated drug score of ≥ 0.50. Wild FCGR2A protein best drug binding pockets are presented in table 6 and Figure: 10 (a) P-0 (Cream in color) with being the highest-scoring pocket with a drug score of 0.71. The second and third highest-scoring pockets were P-1 (blue in color) and P-2 (green in color) with drug scores of approximately 0.71 and 0.66, respectively. While the Mutant FCGR2A protein best drug binding pockets are presented in table 7 and Figure 10 (b) P-0 (Cream in color) with being the highest-scoring pocket with a drug score of 0.68. The second and third highest-scoring pockets were P-1 (blue in color) and P-2 (green in color) with drug scores of approximately 0.66 and 0.65, respectively.

# DISCUSSION

The objective of this study was to investigate the functional and structural characteristics of gene and protein, with the goal of improving our understanding of their implications for therapy. We selected FCGR2A gene and protein for annotation using *in-silico* methods. A systematic analysis was performed to describe the sequence of FCGR2A using the Universal Protein Resource (UniProt), which is a comprehensive resource for protein sequences (24). In this study, the structural analysis and annotation data of RA associated proteins were obtained through the use of UniProt, a comprehensive resource for protein sequences. The physico-chemical properties of proteins, such as their amino acid composition, coefficient, instability, GRAVY, aliphatic index, theoretical pI, atomic composition, and molecular weight, were analyzed to gain a better understanding of their stability, activity, and nature. A high aliphatic index is indicative of thermal stability, with values above 66.5 suggesting that a protein is thermally stable and contains a high proportion of hydrophobic amino acids. To predict these physico-chemical properties of RA associated proteins, we utilized the protparam expassy tool (13) we predicated FCGR2A protein is thermally unstable.

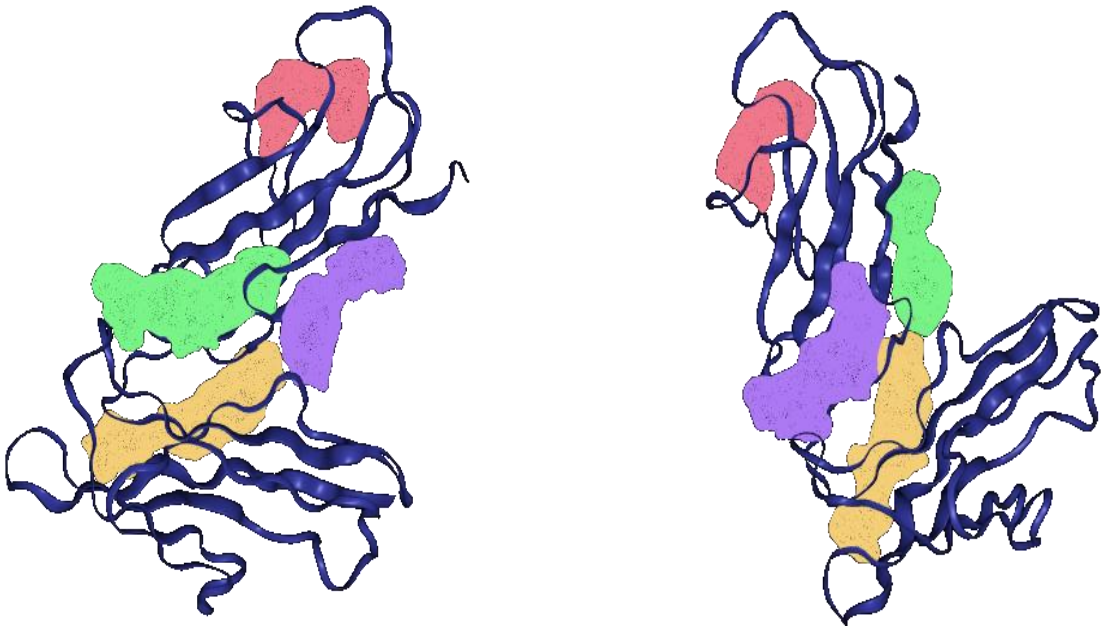


**Fig. 10 (a).** FCGR2A best drug binding pocket **(b).** FCGR2A best drug binding pocket

Post-translation modification (PTM) is a process that occurs after a protein has been synthesized and folded into its three-dimensional structure. PTMs can include a wide range of chemical modifications to the protein, such as phosphorylation, acetylation, glycosylation, ubiquitination, and methylation, among others. These modifications can alter the activity, stability, localization, and interaction partners of the protein, thereby influencing its biological function. PTMs are essential for many cellular processes, including signal transduction, gene expression, and protein degradation. Understanding PTMs is important for studying protein function and developing targeted therapies for various diseases (25). The post-translational modification analysis of the protein revealed the presence of multiple sites of modification including 1 C-mannosylation, glycation, 2 N-linked glycosylation, and serine-containing 1 O-linked glycosylation. Additionally identified a total of 39 phosphorylation sites in the protein of interest. Among these sites, serine residues were found to be the most common, with a total of 20 predicted phosphorylation sites, followed by threonine with 14 predicted sites, and tyrosine with 5 predicted sites. The identified phosphorylation sites were associated with various kinases such as CDK5, PKC, DNAPK, P38MAPK, GSK3, SRC, INSR, PKA, and EGFR. Several studies have estimated that around 60% of the risk for developing RA can be attributed to genetic factors. To better understand how genetic variants associated with RA may affect the activity of candidate proteins and subsequent physiology of effector cells, various attempts have been conducted (4). For FCGR2A, Some studies have shown association of single nucleotide polymorphism (SNP) within

FCGR2A gene with RA. A non-synonymous SNP (rs6671847, rs1801274, rs17400517, and rs6668534) of the FCGR2A gene has been reported to be a RA susceptible polymorphism in RA patients. It express in mononuclear phagocytes, neutrophils, and platelets. Among the FCGR2A polymorphisms studied in Rheumatoid Arthritis (RA), the FCGR2A His167Arg (rs1801274) polymorphism has received particular attention due to its distinct biological functions in comparison to other FCGR genotypes In recent era, there has been a surge in the identification of different genetic variations using computational techniques to analyze data (26). Multiple advanced methods have been developed to align numerous protein sequences and evaluate their conservation across different species. Researchers have conducted a comprehensive alignment of various variant prediction approaches, such as SIFT, Polyphen, CADD, and provean, to study their prediction sensitivities (27). However FCGR2A (rs1801274, His167Arg) Variants are known to have a strong genome significant association with RA, Computational method like SIFT predicate His167Arg has a tolerated effect, polyphen predicted bengin, while the CADD and REVEL method shown likely bengin effect of variant. The variants rs6671847, rs17400517, and rs6668534 were found to have an insignificant impact on protein structure and function based on predictions. These studies highlight that nucleotide variant prediction methods may not always be effective in predicting the pathogenic effects of clinically significant variants. This may be attributed to differences in the datasets used to train these variant prediction programs (28). Consequently, we conducted an additional investigation to assess the influence of the mutation on the structural characteristics of proteins.

It is incredible that a missense mutation can cause the development disorder if it leads to the loss or gain of critical functions due to alterations in the conformation of secondary structures. Secondary structures are fundamental polypeptide structures and the most commonly observed and functionally important polypeptide structure, and these secondary structures fold to form domains, motifs, and tertiary structures (29). The current investigation of the secondary structural analysis of the missense mutations in the His167Arg variants revealed a non-pathogenic effect on RA. The three-dimensional structure of our protein model and superimposition revealed insignificant variations resulting from the presence of mutant amino acid residues (30). The consensus results obtained from the mutation cutoff scanning matrix and site-directed mutator technique indicated a negative Gibbs free energy change for the His167Arg FCFGR2A mutations. This negative change in free energy contributes to the instability of proteins. Various disease-associated mutations can alter the molecular binding energies of molecules, thereby affecting the stability and structural conformation of proteins (31). Our study aimed to investigate the impact of the missense mutation associated with RA in the context of intermolecular interactions. To achieve this, we analyzed the protein-protein interaction network using the STRING database, which predicts both physical (direct) and functional (indirect) relationships. Interactions between proteins are crucial in living systems, as they play significant roles in various cellular functions, including the formation of transient dimers or oligomers, which act as enzymes, transporters, and help maintain the shape of the cell. Identifying protein interactions can improve our understanding of infection mechanisms and facilitate drug development for medication purposes (32).

Our investigation of the protein network of FCGR2A highlights its functional connections with various molecules in the immune system, reinforcing the central role of immune dysfunction in the development of RA. Our molecular docking analysis reveals that missense mutations in RA result in alterations in the interaction with ligand molecules, primarily due to changes in the interacting amino acid residues. To perform molecular docking calculations, we utilized the HADDOCK 2.4 server (33). Our analysis of His167Arg FCGR2A protein revealed significant changes in the interacting amino acid residues and ligand binding energy values, which could affect the interaction with ligand molecules. This SNP may also have an indirect impact on the development of RA by influencing gene expression in areas of linkage disequilibrium. Our findings suggest that His167Arg FCGR2A may play a role in the pathogenesis of RA through its interaction with immune system molecules and influence on gene expression.

The DoG Site Scorer was used to evaluate the druggability potential of the proteins associated with RA, with a focus on identifying potential therapeutic targets. Through prediction of binding pockets, the top

five pockets with the highest druggability scores were selected as the most promising targets for drug development. The drug targets profile for the studied proteins was developed. In the wild FCGR2A protein, we identified four binding pockets, with P-0 having the highest drug score of 0.73 and P-1 having a drug score of approximately 0.67. In the H167A FCGR2A mutant protein, the highest scoring pocket was P-0 with a drug score of 0.68. The second and third highest scoring pockets were P-1 and P-2, with drug scores of approximately 0.66 and 0.65, respectively. These proteins are directly or indirectly involved in the pathophysiology of RA, and utilizing this genomic and proteomic data for subsequent therapeutic options could be useful in treating RA.

# CONCLUSION

The current study aimed to investigate the differences in variant prediction methods for the three best-known RA missense mutations, including pathogenicity prediction, 3D protein structure analysis, and changes in molecular interaction abilities. Various tools, including SIFT, Polyphen, CADD, and REVEL, provided inconsistent results during potential variant analysis of the RA gene, possibly due to differences in the datasets used to train these prediction programs. Therefore, a comprehensive multidirectional approach was used, including gene and protein annotation, sequence based physio-chemical properties, domain analysis, post-translational modification analysis, secondary structure analysis, protein 3D structure modeling, superimposition, protein-protein interaction analysis, protein-protein docking, and developing drugability profiling. These analyses provided a more realistic prediction of the variant effects, but biological assays and protein-protein molecular docking is needed to validate these findings *in-vivo* for RA therapy.

## Acknowledgements:

We would like to express our sincere gratitude to KMU (Khyber Medical University) for their support and resources throughout the duration of this study. We also acknowledge the collaborative spirit and scholarly environment fostered by KMU, which has enriched our research experience and contributed to the advancement of scientific knowledge.

## Conflict of Interest:

Authors have no conflict of interest.

## References:

1. Halushka MK, Fan J-B, Bentley K, Hsie L, Shen N, Weder A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nature genetics. 1999;22(3):239-47.
2. Angelotti F, Parma A, Cafaro G, Capecchi R, Alunno A, Puxeddu I. One year in review 2017: pathogenesis of rheumatoid arthritis. Clin Exp Rheumatol. 2017;35(3):368-78.
3. Yau AC, Holmdahl R. Rheumatoid arthritis: identifying and characterising polymorphisms using rat models. Disease models & mechanisms. 2016;9(10):1111-23.
4. Van der Woude D, Houwing-Duistermaat JJ, Toes RE, Huizinga TW, Thomson W, Worthington J. Quantitative heritability of anti–citrullinated protein antibody–positive and anti–citrullinated protein antibody–negative rheumatoid arthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 2009;60(4):916-23.
5. Sode J, Vogel U, Bank S, Andersen PS, Hetland ML, Locht H, et al. Genetic variations in pattern recognition receptor loci are associated with anti-TNF response in patients with rheumatoid arthritis. PloS one. 2015;10(10):e0139781.
6. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nature genetics. 2015;47(9):1085-90.
7. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nature genetics. 2012;44(5):483-9.

8. Kim K, Bang S-Y, Lee H-S, Bae S-C. Update on the genetic architecture of rheumatoid arthritis. Nature Reviews Rheumatology. 2017;13(1):13-24.
9. Pandey JP. Genomewide association studies and assessment of risk of disease. The New England journal of medicine. 2010;363(21):2076-7.
10. Ravetch JV, Bolland S. Igg fc receptors. Annual review of immunology. 2001;19(1):275-90.
11. Yang Y, Peng L, He C, Xing S, Li D, Jin T, et al. The role of genetic variants in FCGR2A on the risk of rheumatoid arthritis in the Han Chinese population. 2020.
12. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic acids research. 2013;41(W1):W597-W600.
13. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, et al. MFPPI–multi FASTA ProtParam interface. Bioinformation. 2016;12(2):74.
14. Julenius K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. Glycobiology. 2007;17(8):868-76.
15. Kiemer L, Lund O, Brunak S, Blom N. Coronavirus 3CL pro proteinase cleavage sites: Possible relevance to SARS virus pathology. BMC bioinformatics. 2004;5(1):1-9.
16. Johansen MB, Kiemer L, Brunak S. Analysis and prediction of mammalian protein glycation. Glycobiology. 2006;16(9):844-53.
17. Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. 2004.
18. 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. The EMBO journal. 2013;32(10):1478-88.
19. 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;4(6):1633-49.
20. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Engineering Design and Selection. 2004;17(1):107-12.
21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. electrophoresis. 1997;18(15):2714-23.
22. Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics. 2003;50(3):437-50.
23. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez 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.
24. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic acids research. 2000;28(1):45-8.
25. Perez-Vilar J, Randell SH, Boucher RC. C-Mannosylation of MUC5AC and MUC5B Cys subdomains. Glycobiology. 2004;14(4):325-37.
26. Hou L, Zhao H. A review of post-GWAS prioritization approaches. Frontiers in genetics. 2013;4:280.
27. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nature Reviews Genetics. 2011;12(9):628-40.
28. Al-Abbasi FA, Mohammed K, Sadath S, Banaganapalli B, Nasser K, Shaik NA. Computational protein phenotype characterization of IL10RA mutations causative to early onset inflammatory bowel disease (IBD). Frontiers in genetics. 2018;9:146.
29. Khan S, Vihinen M. Spectrum of disease-causing mutations in protein secondary structures. BMC Structural Biology. 2007;7(1):1-18.
30. Miao Z, Cao Y. Quantifying side-chain conformational variations in protein structure. Scientific reports. 2016;6(1):1-10.
31. Hijikata A, Tsuji T, Shionyu M, Shirai T. Decoding disease-causing mechanisms of missense mutations from supramolecular structures. Scientific reports. 2017;7(1):1-8.
32. Rao VS, Srinivas K, Sujini G, Kumar G. Protein-protein interaction detection: methods and analysis. International journal of proteomics. 2014;2014.
33. Van Zundert G, Rodrigues J, Trellet M, Schmitz C, Kastritis P, Karaca E, et al. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. Journal of molecular biology. 2016;428(4):720-5.