

IDENTIFICATION OF HUB GENES INVOLVED IN HUMAN RHEUMATOID ARTHRITIS: AN IN-SILICO APPROACH

Submitted in partial fulfilment of the requirements for the award of Bachelor of
Science degree in Bioinformatics and Data Science

by

SUSHMA T A(40738022)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade “A” by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI – 600 119

MAY – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" Grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.

Phone: 044 - 2450 3150 / 51 / 52 / 54 / 55 Fax: 044 - 2450 2344

www.sathyabama.ac.in



DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **SUSHMA T A (40738022)** who has carried out the project entitled "**IDENTIFICATION OF HUB GENES INVOLVED IN HUMAN RHEUMATOID ARTHRITIS: AN IN-SILICO APPROACH**" under my supervision from January 2023 to May 2023.

Internal Guide

Dr. SWETHA SUNKAR, M.Sc., M.Tech., Ph.D.,

Head of the Department

Dr. JEMMY CHRISTY, M.Tech., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **SUSHMA T A** hereby declare that the Project Report entitled **“IDENTIFICATION OF HUB GENES INVOLVED IN HUMAN RHEUMATOID ARTHRITIS”** done by me under the guidance of **Dr. SWETHA SUNKAR, M.Sc., M.Tech., Ph.D.,** at **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** is submitted in partial fulfilment of the requirements for the award of Bachelor of Science degree in **BIOINFORMATICS AND DATA SCIENCE.**

DATE:9-5-23

SUSHMA T A

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKARAVARTHI, M.E., Ph.D.**, Dean, School of **BIO AND CHEMICAL ENGINEERING** and **Dr. JEMMY CHRISTY, M.Tech., Ph.D.**, Head of the Department, Dept. of **BIOINFORMATICS AND DATA SCIENCE** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. SWETHA SUNKAR, M.Sc., M.Tech., Ph.D.**, for his valuable guidance, suggestions, and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of **BIOINFORMATICS** who were helpful in many ways for the completion of the project.

ABSTRACT

Rheumatoid arthritis (RA) is a chronic, progressive, and disabling autoimmune disease that primarily affects joints. The recapitulation of human RA etiopathogenesis (etiology & pathophysiology) and drug responses are poorly understood, albeit medical literatures have been describing about RA for over 200 years. But the innovation in science and technology has spawned many pertinent and predictive tools. With the help of these tools, our study aims on determining the differentially expressed genes responsible for the development of Rheumatoid arthritis which might assist the identification of the plausible drug targets and the cause of disease management.

A gene expression profile (GSE193193) containing 36107 Differentially Expressed Genes (DEGs) was retrieved from the GEO database and screened using the adjusted p value and log2fc values based on which 3672 DEGs were found to be significant of which 283 genes were up-regulated and 360 genes were down-regulated. Followed by this, the gene enrichment analysis was performed to identify the relating Gene Ontology terms and Pathways. The Protein-Protein Network was constructed with 251 nodes & 647 edges for the up-regulated genes and 232 nodes & 199 edges for the down-regulated genes, then exported to Cytoscape software to predict the hub genes with the help of MCODE and Cytohubba plugins. The identified hub genes are HBB, ALAS2, GATA1, AHSP, HBG1, HBG2, HBD, KLF1, SLC4A1, EPB42, ZMYND10, DNAJC7, HYDIN, LRRC6, FN1, NCAM1, FASLG, CTCF, SMAD4, and STAT1. These hub genes are not just linked to skin and connective tissue diseases like Rheumatoid arthritis but also to a few other diseases. Hence, targeting these genes might possibly aid in the treatment of rheumatoid arthritis. Furthermore, three transcription factors (GATA1, NFKB1, and RELA) and one miRNA (has-mir-27a-3p) were identified as key regulators of hub gene expression.

CONTENTS

CHAPTER NUMBER	TITLE	PAGE NUMBER
	LIST OF TABLES	8
	LIST OF FIGURES	9
	LIST OF ABBREVIATIONS	11
1.	INTRODUCTION	13
	1.1 EPIDEMIOLOGY	13
	1.2 ETIOLOGY	13
	1.3 PATHOGENESIS	14
	1.4 CLINICAL HISTORY	16
	1.5 PHYSICAL EXAMINATION	17
	1.6 EXTRA-ARTICULAR DISEASES	18
	1.7 TREATMENT	20
2.	LITERATURE SURVEY	23
3.	AIM AND SCOPE	26
	3.1 AIM OF THE PROJECT	26
	3.2 SCOPE OF THE PROJECT	26
4.	METHODOLOGY	27
	4.1 RETRIEVAL OF GEO DATASET	27
	4.2 PROCESSING OF DATA FOR DIFFERENTIALLY EXPRESSED GENES	27
	4.3 FUNCTIONAL ENRICHMENT AND PATHWAY ANALYSIS OF DEGs	27
	4.4 CONSTRUCTION OF PROTEIN-PROTEIN INTERACTION NETWORK	28
	4.5 SCREENING OF PPI NETWORK FOR MODULES AND HUB GENES ANALYSIS	28
	4.6 GENE EXPRESSION VALIDATION AND GENE-DISEASE ASSOCIATION ANALYSIS	28

	4.7 IDENTIFICATION OF miRNAs TARGETING THE HUB GENES	29
5.	RESULTS AND DISCUSSION	30
	5.1 DATA ACQUISITION AND ANALYSIS OF DEGs	30
	5.2 FUNCTIONAL ENRICHMENT OF DEGs	30
	5.3 PPI NETWORK CONSTRUCTION	37
	5.4 SCREENING OF PPI FOR MODULES AND HUB GENES ANALYSIS	37
	5.5 GENE EXPRESSION VALIDATION AND DISEASE ASSOCIATION ANALYSIS	43
	5.6 IDENTIFICATION OF miRNAs TARGETING THE HUB GENES	49
	5.7 DISCUSSION	51
6.	SUMMARY AND CONCLUSION	62
7.	REFERENCES	63

LIST OF TABLES

TABLE NUMBER	TABLE NAME	PAGE NUMBER
5.1	LIST OF DIFFERENTIALLY EXPRESSED GENES	31
5.2	GENE-DISEASE ASSOCIATION	44
5.3	(a) FUNCTIONS OF THE HUB GENES FOUND FROM UP-REGULATED NETWORK (b) FUNCTIONS OF THE HUB GENES FOUND FROM DOWN-REGULATED NETWORK	45 46
5.4	GENES ASSOCIATION WITH OTHER DISEASES	48

LIST OF FIGURES

FIGURE NUMBER	FIGURE NAME	PAGE NUMBER
5.1	ENRICHMENT ANALYSIS OF UP-REGULATED GENES PERFORMED BY EVALUATING THE GENE ONTOLOGY TERMS	32
5.2	ENRICHMENT ANALYSIS OF UP-REGULATED GENES PERFORMED BY EVALUATING THE PATHWAY ASSOCIATIONS	33
5.3	ENRICHMENT ANALYSIS OF DOWN-REGULATED GENES PERFORMED BY EVALUATING THE GENE ONTOLOGY TERMS	34
5.4	ENRICHMENT ANALYSIS OF DOWN-REGULATED GENES PERFORMED BY EVALUATING THE PATHWAY ASSOCIATIONS	35
5.5	PPI NETWORK CONSTRUCTED USING STRING DATABASE	37
5.6	VISUALIZATION OF PPI NETWORK USING CYTOSCAPE	38
5.7	MODULE ANALYSIS USING MCODE	39

5.8	COMMON HUB GENES IN UP-REGULATED NETWORK FROM CYTOHUBBA ON MCODE	40
5.9	COMMON HUB GENES IN DOWN-REGULATED NETWORK FROM CYTOHUBBA ON MCODE	41
5.10	COMMON HUB GENES IN UP-REGULATED NETWORK FROM CYTOHUBBA ON ENTIRE NETWORK	42
5.11	COMMON HUB GENES IN UP-REGULATED NETWORK FROM CYTOHUBBA ON ENTIRE NETWORK	43
5.12	NETWORK OF REGULATORY INTERACTIONS BETWEEN miRNAs, TRANSCRIPTION FACTORS, AND HUB GENES	50
5.13	FILTERED NETWORK OF miRNAs AND TRANSCRIPTION FACTORS TARGETING HUB GENES	51

LIST OF ABBREVIATIONS

RA	Rheumatoid Arthritis
ASR	Age Standardized Rate
NHS	National Health Survey
RF	Rheumatoid Factors
AMPA	Anti-Modified Protein Antibodies
SE	Shared Epitope
MMP	Matrix Metalloproteinases
TNF	Tumour Necrosis Factor
IL	Interleukin
RANKL	Receptor Activator of Nuclear factor- Kappa B Ligand
MCP	Metacarpophalangeal
MTP	Metatarsophalangeal
Anti-CCP	Anti-Cyclic Citrullinated Peptide Antibodies
Abs	
ESR	Erythrocyte Sedimentation Rate
ILD	Interstitial Lung Disease
Anti-CarP	Anti-Carbamylated Protein
Anti-MAA	Anti-Malondialdehyde Antibodies
ACR	American College Rheumatology
SDAI	Simplified Disease Activity Index
CDAI	Clinical Disease Activity Index
DMARD	Disease Modifying Antirheumatic Medications
JAK	Janus Kinases
NSAID	Non-Steroidal Anti-Inflammatory Drugs
GEO	Gene Expression Omnibus
DEG	Differentially Expressed Genes

P value	Probability Value
DAVID	Database for Annotation, Visualization, and Integrated Discovery
GO	Gene Ontology
MF	Molecular Function
BP	Biological Process
CC	Cellular Component
PPI	Protein-Protein Interaction
MCODE	Molecular Complex Detection
TF	Transcription Factor
miRNA	microRNA

CHAPTER – 1

1. INTRODUCTION:

1.1 EPIDEMIOLOGY:

According to the Global Burden of disease 2019 study, the worldwide prevalence of Rheumatoid arthritis is increased from 567,462.89 to 1,074,390.80 cases with an ASR (age-standardized rate) of 13/100,000 patients between 1990 and 2019. It is estimated to affect approximately 0.5% to 1.0 % of the population in developed countries and 0.6 % of the population in United States (Gabriel S.E. and Michaud K., 2009). Based on the self-reported data of Australia and from 2017-2018 National Health Survey (NHS), Australia has reported the highest prevalence of RA worldwide which is 1.9%. It is reported that the incidence of RA in women is two to three times greater than in man, that is, 1 in 12 women and 1 in 20 men develop an inflammatory autoimmune rheumatic disease during their lifetime (Crowson et al., 2011). Although rheumatoid arthritis may occur at any age, most commonly the patients are affected between 25 to 50 years of age (Cassotta et al., 2019). Ted R. Mikuls's study on the incidence of rheumatoid arthritis found that about one-fourth to one-half of the patients with RA become unable to work within 10 to 20 years of follow-up (Mikuls, 2010).

1.2 ETIOLOGY:

The etiology of RA is insufficiently understood, but it is presumed that RA can be caused from the combination of genetic risk factors and autoantibodies. The aberrations in adaptive immune system causes the development of autoantibodies namely Rheumatoid Factors (RF) and Anti-Modified Protein Antibodies (AMPA). AMPA comprises of antibodies that works against various modifications such as

citrullination (ACPA), Carbamylation (aCarP) and Acetylation (AAPA) (HU Scherer et al., 2020). Patients diagnosed with seropositive rheumatoid arthritis are likely to inherit approximately 40% to 65% of the disease from their ancestors, while those diagnosed with seronegative rheumatoid arthritis inherit approximately 20% of the disease (Chauhan K. et al., 2022) (Klodzinski, L. and Wislowska, M., 2018). According to the Arthritis Foundation, 80% of the patients diagnosed with RA comprises of Rheumatoid Factor and 60% to 70% of the patients with RA comprises the antibodies of Cyclic Citrullinated Peptide and this does not mean that the person has developed RA, as RF also indicates other conditions (Arthritis Foundation, 2021). The hypothesis of Shared Epitope (SE) describes that class II molecules are involved directly in the etiopathogenesis of autoimmune diseases (Gregersen, P.K. et al., 1987). Inheritance of HLA-DRB1 alleles of MHC Class- II are the strongest genetic risk factors that increases the susceptibility to RA (Cornelia M. et al., 1992). HLA-DRB1 alleles namely HLA-DRB1*04, HLA-DRB1*01 and HLA-DRB1*10 codes for a sequence of five amino acids called “shared epitope”, which is carried by 80% of ACPA positive RA patients (Derksen V.F.A.M. et al., 2017) (du Teil Espina, M. et al., 2019). When the carriers of HLA-DR SE genes are exposed to environmental conditions such as smoking, the risk of developing RA is higher, as these genes can trigger the RA-specific immune response against citrullinated proteins (Klareskog, L. et al., 2006).

1.3 PATHOGENESIS:

Autoantibodies and autoimmunity are found to be developed even before the genesis of clinical phase of RA. Despite having autoantibodies, many patients do not develop any obvious illness. In some patients, autoimmunity transforms into the immune-mediated inflammation that largely occurs in the synovium (Chauhan K. et al., 2022). Patients with rheumatoid arthritis have an accumulation of autoreactive T and B lymphocytes in their synovial tissues. Immunologically, T cells are tolerant to autoantigens, but when self-tolerance is compromised, autoreactive T cells become activated and prompt the B cells to cause the generation of autoantibodies. Autoantibodies combines with antigen to produce immune complexes that are deposited in the tissues. Angiogenesis or vasodilation,

synoviocyte proliferation, and lymphocyte accumulation are characteristics of tissues with synovitis (Tanaka, Y., 2020). Synovial inflammation causes pannus to develop, which damages the neighboring articular structures. Generally, the synovium contains a cellular structure with a delicate interior lining. In rheumatoid arthritis, the macrophages, plasma cells, dendritic cells, lymphocytes, and immune complexes enter the synovium and form distinct lymphoid masses with germinal centers. The synoviocytes in a healthy synovium come from mesenchymal cells, but in RA, the number of synoviocytes that resemble macrophages and fibroblasts is rising as the membrane enlarges. These cells produce a variety of cytokines, chemokines, adhesion molecules, matrix metalloproteinases (MMPs), tissue inhibitors of metalloproteinases and aggrecanases that breaks down the extracellular matrix and the articular structures (Alam, J. et al., 2017). The two pro-inflammatory cytokines, TNF- α (Tumor Necrosis Factor – alpha) and IL-1 (Interleukin-1) are the key cytokines that plays a crucial role in the development of chronic joint inflammation and the corresponding erosive alterations in bone and cartilage. Studies using animal models have tremendously aided in this identification. In the initial studies, recombinant cytokines were administered directly into the knee joints of rabbits and rodents to assess their potential for arthritogenicity. This was the first circumstantial proof that TNF- α was an inflammatory mediator and that IL-1 was a key cytokine involved in the development of arthritis and cartilage damage. Although TNF- α was not particularly damaging itself, it might synergistically increase IL-1's destructiveness (Erik Lubberts and Wim B. van den Berg., 2000-2013). Osteoclast maturation and activation are induced by receptor activator of nuclear factor-kappa B ligand (RANKL), which is expressed by lymphocytes and synoviocytes. Up until they come in touch with the bones, the inflammatory granulation tissues comprising of proliferative and stratified synoviocytes grow. Joint deterioration happens mostly at the point of their contact, which occurs as a result of multinucleated osteoclasts destroying and absorbing the bones (Tanaka, Y., 2020).

1.4 CLINICAL HISTORY:

Rheumatoid arthritis is a systemic, autoimmune disease in which the body's immune system mistakenly attacks the joints, creating inflammation that causes the tissues lining the inside of the joints to thicken, resulting in swelling and pain in and around the joints (Arthritis Foundation, 2019). RA generally starts affecting small peripheral joints, if it is left untreated it progresses to involve proximal joints and the commonly affected joints include, joints of hands, feet, wrists, elbows, knees, and ankle (Bullock, J. et al., 2018). Usually, the joint pain in RA is symmetrical and polyarticular but it can also be asymmetrical, oligoarticular, or monoarticular at the onset (Sparks, J.A., 2019). Patients diagnosed with Palindromic rheumatism (a relapsing arthritis syndrome) experience episodic sudden onset of attacks in joints or periarticular tissues, affected over hours to days. Two-third of patients with Palindromic rheumatism, developed chronic arthritis (Essi Koskinen et al., 2009). The Characteristic symptoms of RA involve morning stiffness of the affected joints which occurs for more than 30 minutes and because of this stiffness at the onset of the disease, patients face difficulties in moving their fingers upon awakening (Tanaka, Y., 2020). Other non-specific four crucial extra-articular symptoms that impact the lives of patients diagnosed with Rheumatoid arthritis are pain, depression & anxiety, fatigue & rheumatoid cachexia, and these symptoms are frequently observed in them. Studies have reported that one in four patients experience high levels of fatigue (Scott, I.C. et al., 2018). The pattern of symptoms may increase and decrease within a day or even from day to day. Incidences of flares is a common symptom for patients with RA and are experienced with increased joint pain and deteriorations in inflammation (McWilliams DF et al., 2019). At times, fever occurs periodically and is always almost low grade (37.5 °C - 38.3°C), which is caused by the inflammatory mediators that function as endogenous pyrogens and drives the inflammatory process (Galloway J and Cope AP., 2015).

1.5 PHYSICAL EXAMINATION:

Physical Examination could reveal articular and soft tissue inflammation with palpable discomfort caused by synovitis. Large or medium joints might have visible effusions that can be treated with arthrocentesis. In the small joints of patients with obesity or concurrent fibromyalgia, mild joint inflammation can be inconspicuous and challenging to diagnose (Sparks, J.A., 2019). Due to the occurrences of pain and swelling in the joints, the range of movement becomes decreased for the patients. The squeeze test is a procedure in which the knuckles of the hands or feet are squeezed across the MCP (Metacarpophalangeal) or MTP (Metatarsophalangeal) joints to examine the level of discomfort, and it is a frequently utilized means of testing. The squeeze test is a useful tool to contribute to the clinical picture, but it is insufficient on its own to diagnose the early arthritis, therefore healthcare professionals should feel confident identifying the swollen joints even in the absence of pain or tenderness. Once RA is clinically suspected, it is important to perform the laboratory tests to determine the disease's severity and to help confirm the diagnosis. These tests include rheumatoid factor (RF), anti-cyclic citrullinated peptide antibodies (anti-CCP Abs), inflammatory markers including the erythrocyte sedimentation rate (ESR) and CRP, and the fundamental laboratory work like complete blood count and full metabolic panel. The antibodies RF and anti-CCP Abs determine whether a patient has seropositive RA. Antibodies called RFs attach to the Fc region of immunoglobulin G, and as a result, they are seen in conditions besides RA. According to reports, RF's overall sensitivity for diagnosing RA is 69%, and in young, healthy individuals, the specificity can reach as high as 85%. The more recent anti-CCP test has an astounding 97% specificity for RA and according to pooled meta-analyses, a sensitivity is of roughly 67%. According to studies, RF testing alone is not as useful in patients with early arthritis as RF testing in combination with anti-CCP antibody testing, which has a much higher specificity and positive predictive value. It is essential to remember that 15% - 25% of RA cases can be seronegative, as indicated by the absence of RF and anti-CCP Abs. Seronegative patients are typically less prone to develop erosive joint disease than seropositive patients,

while they may experience substantial joint pain and swelling (Littlejohn, E.A. and Monrad, S.U., 2018). Physical examination might reveal the decreased grip strength in cases of multiple joint involvement. In the advanced chronic disease, the historical physical findings of ulnar deviation, metacarpophalangeal joint subluxation, swan neck deformity, boutonniere deformity, and the “bowstring” sign are observed (Chauhan K. et al., 2022).

1.6 EXTRA-ARTICULAR DISEASES:

Extra-articular involvement in RA is prevalent, despite the fact that it is categorized as an inflammatory joint disease of symmetric synovial joints (Littlejohn, E.A. and Monrad, S.U., 2018). Due to the systemic nature of RA, either extra-articular organ symptoms or adverse effects from therapy, or both, may occur (Sparks, J.A., 2019). About 40% of RA patients experience extra-articular symptoms at some point early on or later in the disease. Once after the onset of RA, extra-articular symptoms might appear at any age. Each country has a different average frequency of extra-articular RA symptoms (Cijocar, M. et al., 2010). Patients with severe, active disease and those who have positive rheumatoid factor / anti-CCP Abs and/or HLA-DR4 are more likely to experience extra-articular organ involvement in RA. The extra-articular symptoms include skin, eye, heart, lung, renal, neurological, and gastrointestinal organ systems, as well as the involvement of tendons, ligaments, and fascia (Littlejohn, E.A. and Monrad, S.U., 2018). The most prevalent extra-articular symptoms of RA are rheumatoid nodules. In addition to the usual pressure points like olecranon, the rheumatoid nodules can also develop over the joints in hands and feet, over the patella, and on the Achilles tendons (Chauhan K. et al., 2022). Keratoconjunctivitis, an inflammation of the cornea and conjunctiva, is one of the most frequent ocular symptoms. It can cause burning, itchy eyes, pressure behind the eye, or a gritty sensation inside the eye. A less prevalent but more severe condition called scleritis, causes the sclera to become extremely painful and inflamed (Littlejohn, E.A. and Monrad, S.U., 2018). Pulmonary involvement in RA is common but not often clinically identified. The majority of cases of pleural disease are asymptomatic and autopsy studies found that 50% of cases had pleural involvement, with just 10% clinically being identified.

Exudates with mixed cell numbers and high protein concentrations typically characterize pleural effusions. Although highly selective, multinucleated giant cells are only observed in less than 50% of instances. The condition frequently coexists with interstitial lung disease and exudative pericarditis (Cijocar, M. et al., 2010). Around 5 to 16% of RA patients develop interstitial lung disease (ILD), which is linked to elevated mortality and autoantibodies unique to the condition. Before experiencing joint inflammation and having a positive RF and ACPA, certain patients may exhibit ILD. Anti-carbamylated protein antibodies (anti-CarP), anti-malondialdehyde antibodies (anti-MAA), and ACPA are all linked to RA-associated ILD. RA patients with ILD have the MUC5B promoter variant rs35705950. This gene encodes mucin 5B and this variant is a gain of function. The odd ratio for RA with ILD is 3.8 (95% CI 2.8-5.2) when compared to unaffected controls, and it is 3.1 (95% CI 1.8-5.4) when compared to RA without ILD (Chauhan K. et al., 2022). Atherosclerosis is the main cardiac concern in these patients due to the systemic inflammatory nature of RA. According to studies, people with RA have a higher risk of cardiovascular death, ischemic heart disease, and heart failure when compared to controls who are age-matched and sex-matched to them. According to the same study, there was more evidence of inflammation and an increased percentage of unstable plaques in the coronary artery tissue of autopsied RA patients. Therefore, indications of claudication, chest discomfort, or an uncommon anginal analogue must be carefully taken into account. Osteoporosis in RA results from a concurrent use of glucocorticoids, systemic and local inflammation, immobility caused by joint injury and discomfort, and patient immobility. These, together with primary osteoporosis risk factors, demand dual-energy X-ray absorptiometry screening and bone health management. In RA, hematologic abnormalities are typical, and medicinal treatments and disease processes both have the potential to decrease bone marrow. In terms of clinical manifestations, this may take the form of anemia, leukopenia, neutropenia, eosinophilia, thrombocytopenia, or thrombocytosis. The most prevalent anomaly is anemia of chronic disease, which can occur concurrently with iron deficiency anemia and tends to improve with excellent disease control. Asymmetric sensory and/or motor peripheral neuropathy known as mononeuritis multiplex involves impairment to at least two distinct, seemingly unrelated nerve regions. Although it can occur in

vasculitic processes like RA, mononeuritis multiplex is frequently seen in diabetes (Littlejohn, E.A. and Monrad, S.U., 2018).

1.7 TREATMENT:

An essential tenet of the RA management is a notion that disease-modifying therapy should begin as soon as possible following the diagnosis, with the aim of achieving minimal disease activity or total remission. In accordance with the ACR's (American College of Rheumatology) current definition of early RA, treatment should begin when the disease has not yet progressed for six months. Delays of even a few months can change the long-term outcomes in the treatment of RA, which has been found to avoid persistent functional impairment and irreversible structural damage. Early recognition and referral of RA are essential for attaining the best results because primary care professionals are frequently the initial point of contact for patients presenting with new arthritis. However, there are essential therapeutic techniques that primary care clinicians can implement while waiting for specialised examination (Littlejohn, E.A. and Monrad, S.U., 2018). An extensive evaluation of the disease activity, imaging results (such as those from radiography), complications, and comorbidities should be used to select the best therapeutic approaches. The SDAI, CDAI, and DAS28 are three common composite objective indices used to assess disease activity (Tanaka, Y., 2020). The Disease Activity Score 28 (DAS28) is a frequently used measure that includes tender and swollen joints enumerated at 28 places (upper extremities and knees), the patient's self-assessment of global status, and acute-phase reactant measurement (ESR or CRP). An online calculator (www.das-score.nl/dasculators.html) that accepts these raw numbers can be used to calculate the final DAS28 score. Similar to the DAS28 but simpler to construct, the Simplified Disease Activity Index (SDAI) might be more suitable for real-time usage if laboratory findings are available. The Clinical Disease Activity Index does not involve laboratory tests or a calculator; instead, it relies on patient / physician assessments of their overall health and the presence of painful or swollen joints (Sparks, J.A., 2019). The aim of treatment is to achieve remission, which is characterised as a clinical state in which there is no further progression of joint degeneration or dysfunction. Remission criteria have been statistically constructed

using Boolean remission and numerical targets, such as an SDAI score ≤ 3.3 and a CDAI score ≤ 2.8 (Tanaka, Y., 2020). Methotrexate, hydroxychloroquine, sulfasalazine, and leflunomide are examples of disease-modifying antirheumatic medications (DMARDs) that are commonly used in the treatment of RA. Inhibitors of anti-TNF-alpha include etanercept, infliximab, adalimumab, and certolizumab, golimumab, and certolizumab pegol. Interleukin (IL) 6 inhibitors like tocilizumab and sarilumab, T-cell costimulation inhibitors like abatacept (CTLA4-Ig), and the anti-CD20 B-cell depleting monoclonal antibody rituximab are examples of other biological DMARDs. Janus Kinases (JAK) inhibitors like tofacitinib, baricitinib, and upadacitinib are examples of targeted synthetic DMARDs. In patients with a major active infection, DMARD therapy, including biologic medicines and targeted synthetic drugs (tofacitinib), should be temporarily suspended. Once the infection has subsided and the antibiotic treatment is finished, they can be resumed. Patients with liver diseases should avoid methotrexate. Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) are frequently used to treat joint inflammation and pain symptoms; however, they do not have any effects on the course of disease. There are roughly 20 such drugs that are all effective at full doses (depending on the nation you are in). the toxicities and side effects vary to some extent. NSAIDs differ in their effects on the cardiovascular system, however there is a class Black Box warning for cardiovascular disease. Additionally, NSAIDs carry a risk of gastrointestinal, renal, and hematologic damage. The patient, their comorbidities, and the provider's experience with the particular NSAID all play a role in the decision of which NSAID to use. Patients with RA frequently utilize corticosteroids. There are numerous circumstances in which corticosteroids ought to be taken into account. Corticosteroids can be administered as a bridge therapy in a new patient with very active RA while DMARD medication is being initiated. According to studies, administering corticosteroids early in the course of RA patients led to better outcomes and had disease-modifying effects, including radiographic progression. Another study found no long-term advantages of corticosteroid bridge treatment. Mild RA flare-ups can be treated with brief courses of corticosteroids. Single-joint flares may benefit from the administration of intra-articular corticosteroids. Prednisone 2.5-7.5 mg (low-dose corticosteroid) daily is required for about 50% of RA patients in order to keep their conditions under control. Numerous toxicities, such as weight gain, osteoporosis, and an elevated risk of

infections, are linked to long-term corticosteroids. Rituximab is a biological DMARD that can be used to treat RA patients who have uncontrolled RA and have not responded to TNF inhibitors. When administered intravenously, rituximab depletes CD20+ B-cells and reduces the immunological response to vaccinations in its target patients. Rituximab responses are improved in seropositive patients who are concurrently receiving methotrexate. Abatacept binds to CD80 and CD86 to prevent T-cell activation. It is given either as a weekly subcutaneous injection or as a monthly intravenous infusion. Abatacept medication, with established efficacy from 6 months to 5 years of therapy, benefits patients with uncontrolled RA who have exhibited an unsatisfactory response to methotrexate and TNF inhibitory therapy (Chauhan K. et al., 2022).

CHAPTER – 2

2. LITERATURE SURVEY:

The article by Tanaka Y., 2020, provides an overview of Rheumatoid arthritis, a chronic autoimmune disease that affects the joints and lead to significant disability if left untreated. The author describes the pathophysiology of rheumatoid arthritis, including the role of various immune cells and cytokines in the disease process. The article also discusses the clinical presentation of rheumatoid arthritis, including common symptoms such as joint pain, swelling, and stiffness. The diagnosis and treatment of rheumatoid arthritis are also covered, including the use of disease-modifying antirheumatic drugs (DMARDs) and biologic agents. Finally, the article highlights the importance of early diagnosis and treatment in preventing joint damage and improving patient outcomes.

Chauhan K. et al., 2022, published a journal on Rheumatoid arthritis, which explains how anti-citrullinated protein antibodies & rheumatoid factor are used to assess rheumatoid arthritis. This Journal summarizes the Pathophysiology of rheumatoid arthritis and identifies the comorbid conditions associated with rheumatoid arthritis. The importance of enhancing interprofessional care coordination in order to improve the outcomes for people with rheumatoid arthritis has been discussed in this journal.

Adam McDermaid et al., 2019, published an article about the Interpretation of differential gene expression results of RNA-seq data. This article examined DGE results analysis from a functional point of view for various visualizations. It also includes the Visualization of Differential Gene Expression Results using R/Bioconductor package, which creates information-rich visualizations for the interpretation of DGE results from the three widely used tools Cuffdiff, DESeq2, and edgeR. Five real-world data sets, including one each for a human, a *Malus domestica*, and three *Vitis riparia*, are used to test the implemented functionalities.

The study conducted by Ma J. et al., 2021, aimed to identify the key candidate genes and pathways associated with laryngeal squamous cell carcinoma (LSCC) using bioinformatics analysis. The researchers performed a differential gene expression analysis between LSCC tissues and normal tissues, and identified differentially expressed genes (DEGs). They then conducted a Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analysis to identify the biological processes and pathways in which the DEGs were involved. A protein-protein network was constructed using STRING and Cytoscape, and module analysis was constructed. The researchers identified the hub genes using CytoHubba plugin of Cytoscape based on maximal clique centrality. UALCAN and GEPIA data were used to confirm the gene expression results. The connection between the hub genes and the overall survival of LSCC patients was studied using The Cancer Genome Atlas. Finally, the researchers verified the key genes using western blot, qRT-PCR, CCK-8, wound healing, and transwell assays.

Udhaya Kumar S et al., 2020, has published an article that explains about the Analysis of Differentially Expressed Genes and Molecular Pathways in Familial Hypercholesterolemia Involved in Atherosclerosis: A Systematic and Bioinformatics Approach. This work has discovered the core genes and dysregulated molecular pathways that are differentially regulated in FH, as well as the relevant genetic variables and underlying mechanism that increases the risk of atherosclerosis in FH patients. Here, the GEO database's Affymetrix microarray dataset and the GEO2R statistical tool were used to identify the DEGs in WBCs from 5 heterozygous FH patients and 5 healthy controls. The interaction of the DEGs was discovered using the STRING tool and visualised using the Cytoscape software. The gene cluster in the interaction networks was determined using MCODE. The detected DEGs were subjected to functional annotation, such as gene ontology and enriched molecular pathway analysis, using DAVID website and CluePedia.

The article by Yang R, et al., 2020, describes a bioinformatic analysis of differentially expressed genes in tumor and paracancerous tissues of patients with

lung adenocarcinoma. The objective of this study was to identify the genes and signaling pathways associated with lung adenocarcinoma through the analysis of gene expression profiles of lung adenocarcinoma tumor and adjacent tissues using bioinformatics tools. To achieve this, three microarray datasets from the Gene Expression Omnibus database were downloaded and integrated to identify common differentially expressed genes. The protein-protein interaction network of lung adenocarcinoma was constructed using the STRING database to identify gene modules and core genes. The core genes were further validated and analyzed using GEPIA and Kaplan-Meier plotter online tools.

The research work conducted by Modi, A. et al., 2020, describes an in-silico analysis of genes and microRNAs that are associated with lymph node metastasis in invasive breast carcinoma. The study used Gene Expression Omnibus datasets to compare gene expression in lymph node-positive breast cancer patients to nodal-negative and healthy tissue. The authors conducted functional enrichment analysis of the differentially expressed genes using the DAVID tool. They constructed a PPI network using STRING and visualized it using Cytoscape to identify candidate hub genes. The authors analyzed the expression of these candidate genes for survival (OS) using Gene Expression Profiling Interactive Analysis (GEPIA). Finally, the study also analyzed potential target microRNAs and transcription factors associated with these genes using miRNet.

The article written by Qiu X et al., 2021, describes a bioinformatics study aimed at identifying key genes and microRNAs associated with idiopathic pulmonary arterial hypertension (IPAH). To achieve this, the researchers analyzed gene expression data from IPAH patients and healthy controls using bioinformatics tools. They identified differentially expressed genes and used a weighted gene co-expression network analysis to find modules of genes strongly correlated with IPAH. They then conducted functional enrichment analysis and constructed PPI networks to identify hub genes. Finally, they identified miRNAs that target the hub genes and identified the most important miRNAs associated with IPAH.

CHAPTER – 3

3. AIM AND SCOPE OF THE PROJECT:

3.1 AIM OF THE PROJECT:

To identify the novel genes that are associated with the progression of Rheumatoid arthritis by using datasets. This is achieved by:

- retrieving and processing a previously available dataset on RA for identifying the differentially expressed genes.
- analyzing the functional enrichment of the identified genes
- constructing PPI networks with the genes
- screening the PPI networks to identify the cluster modules and hub genes

3.2 SCOPE OF THE PROJECT:

The discovered hub genes might offer breakthrough ideas for developing diagnostics as well as a platform for generating Rheumatoid arthritis therapies.

CHAPTER – 4

4. METHODOLOGY:

4.1 RETRIEVAL OF GEO DATASET

The dataset analyzed in this present study is freely available and was retrieved from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The keyword 'Rheumatoid arthritis' was used to search in GEO datasets and after a careful review, the gene expression profile GSE193193 was selected from the resulting datasets (Fan D et al., 2022). The dataset chosen is based on the platform GPL24676, Illumina NovaSeq 6000 (Homo sapiens).

4.2 PROCESSING OF DATA FOR DIFFERENTIALLY EXPRESSED GENES:

The selected dataset is a curated dataset that comprises the Differentially Expressed Genes (DEGS). From the total set of DEGs, significant genes were filtered based on the adjusted probability value (p value) which is less than 0.05. From the significant genes, the up-regulated and down-regulated genes were detected and categorized based on the cut-off criteria of log fold change ($\log_2fc$) values such as $\log_2fc > 1.5$ and < -1.5 respectively. A similar method has been utilized in a study to analyse the differentially expressed genes and molecular pathways in Familial Hypercholesterolemia involved in Atherosclerosis (Udhaya Kumar, S. et al., 2020).

4.3 FUNCTIONAL ENRICHMENT AND PATHWAY ANALYSIS OF DEGs:

The detected up-regulated and down-regulated DEGs were further analyzed for their functional and enrichment analysis using a Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool (Huang et al., 2009). The DAVID tool shows Gene Ontology (GO) that represents enriched genes in three different terms namely Molecular Function (MF), Biological Process (BP), Cellular Component (CC) and pathway related information of genes from different databases such as KEGG, Reactome, etc. The pathway analysis shows the

involvement of DEGs in different pathways dealing with cancer related pathways, diseases, drugs, and chemical substance. The cut-off criteria were used for p value as $p < 0.05$ and gene count > 2 .

4.4 CONSTRUCTION OF PROTEIN-PROTEIN INTERACTION NETWORK (PPI):

The PPI networks of the DEGs were predicted and constructed using a search tool (STRING, version 11.5; <https://string-db.org/>) for the retrieval of interacting genes/proteins (Szkarczyk et al., 2015). Based on the STRING tool, the PPI networks of the DEGs (up-regulated and down-regulated genes) were constructed separately with the threshold of confidence score > 0.4 .

4.5 SCREENING OF PPI NETWORK FOR MODULES AND HUB GENES

ANALYSIS:

Cytoscape version 3.9.1 (<https://cytoscape.org/>), an open-source software, was used for visualizing the networks and integrated networks to find their molecular interactions and associated biological pathways. Subsequently, the PPI network retrieved from STRING database was exported to Cytoscape and visualized for further analysis (Shannon et al., 2003). The Molecular Complex Detection (MCODE) plug-in of Cytoscape was used to determine the densely connected regions using the most significant cluster module from the PPI network. CytoHubba, another plug-in of Cytoscape, was used to predict the hub genes in the network based on the 12 topological algorithms of CytoHubba for ranking the nodes in a PPI network by network characteristics.

4.6 GENE EXPRESSION VALIDATION AND GENE-DISEASE ASSOCIATION

ANALYSIS:

Online tools such as GeneMANIA, DAVID, UniProt & NCBI and health information resources like MedlinePlus Genetics were used to identify the functions of the predicted hub genes and their role in other diseases. DisGeNET (<https://www.disgenet.org/home/>), an online platform that contains publicly

available collections of genes and variants associated to human diseases was used to identify the relationship between hub genes and Rheumatoid arthritis.

4.7 IDENTIFICATION OF miRNAs TARGETING THE HUB GENES:

miRNet 2.0 (<https://www.mirnet.ca/miRNet/home.xhtml>), a miRNA-centric network visual analytics platform was used to discover the miRNAs and transcription factors targeting the hub genes (*Fan et al., 2016*). The relationship between transcription factors and the hub genes were analyzed using TRRUST database in miRNet. Network tools such as Degree filter was used to identify the nodes with higher degrees that act as important hubs in the network by reducing the less important nodes.

CHAPTER - 5

5. RESULTS AND DISCUSSION:

5.1 DATA ACQUISITION AND ANALYSIS OF DEGs:

The GSE193193 dataset containing the differentially expressed genes was obtained from the GEO database. Out of 36107 DEGs found in the dataset, 3672 genes were significant based on the p-value <0.05 and out of which 283 genes were significantly up-regulated and 360 genes were significantly down-regulated based on the log2fc values >1.5 and <-1.5 respectively. The first 20 genes which are up-regulated and down-regulated with p-value and log2fc values are provided in Table 5.1.

5.2 FUNCTIONAL ENRICHMENT OF DEGs:

DAVID online software was used for performing the functional annotation of DEGs. The Gene ontology terms including Biological Process (BP), Molecular Function (MF), Cellular Component (CC) and the pathways such as KEGG and REACTOME pathways were evaluated for both up-regulated and down-regulated genes separately. By analysing BP, it was found that the up-regulated genes were enriched in innate immune response, negative regulation of apoptotic process and cell adhesion. Among the MF categories, the up-regulated genes were significantly enriched in identical protein binding and zinc ion binding. Within the CC categories, the up-regulated genes were especially linked with plasma membrane, cytosol, and extracellular exosome (Fig 5.1). In addition, the KEGG pathway enrichment analysis revealed that the up-regulated genes were mainly involved in transcriptional mis-regulation in cancer and phagosome. The REACTOME pathway analysis revealed that the up-regulated genes were mostly involved in immune system and innate immune system (Fig 5.2). Similarly, Gene ontology and pathway analysis were performed for the down-regulated genes. By analysing BP, it was found that the down-regulated genes were enriched in cell adhesion, nervous system development and homophilic cell adhesion via plasma membrane

adhesion molecules. Among the MF categories, down-regulated genes were significantly enriched in calcium ion binding. Within the CC categories, the down-regulated genes were linked with plasma membrane and integral component of membrane (Fig 5.3). In addition, the KEGG pathway enrichment analysis revealed that the down-regulated genes were involved in antigen processing & presentation and natural killer cell mediated cytotoxicity. The REACTOME pathway analysis revealed that the down-regulated genes were involved in signal transduction and developmental biology (Fig 5.4).

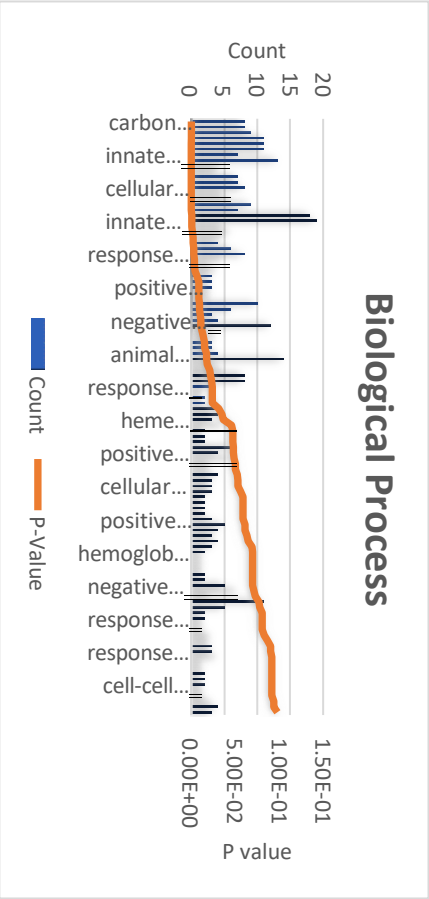
Table 1.1: List of differentially expressed genes

(a) List of up-regulated genes:

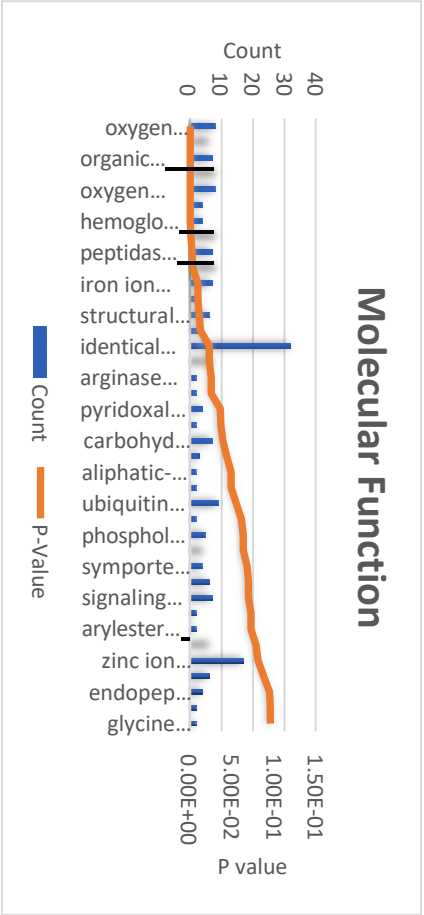
gene_name	log2(fc)	pval	regulation
RNU11	4.58	0.00	up
HBD	9.56	0.00	up
CA1	7.78	0.00	up
ALAS2	7.69	0.00	up
HBG1	6.10	0.00	up
HBA1	6.15	0.00	up
SAMD14	2.32	0.00	up
SLC4A1	6.37	0.00	up
HBB	6.68	0.00	up
HBA2	5.78	0.00	up
SELENBP1	6.60	0.00	up
OSBP2	4.49	0.00	up
AHSP	9.37	0.00	up
MYL4	8.00	0.00	up
SNCA	4.35	0.00	up
HBG2	6.40	0.00	up
KRT1	6.32	0.00	up
RUNDC3A	6.88	0.00	up
SLC6A8	5.34	0.00	up

(b) List of down-regulated genes:

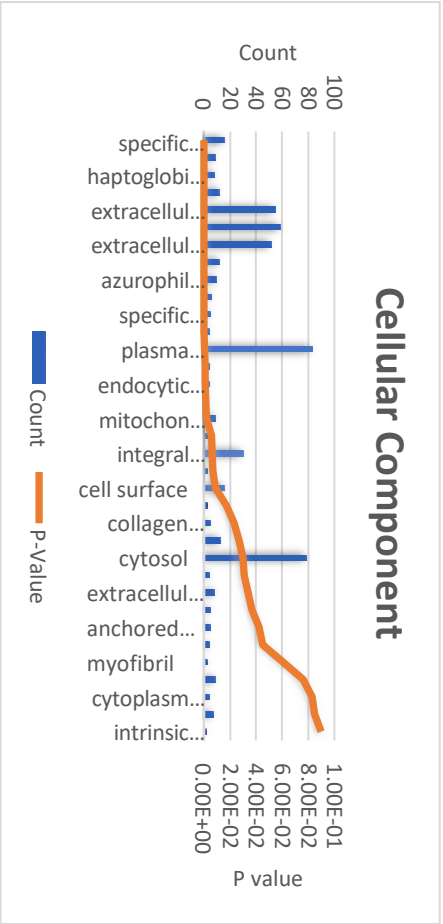
gene_name	log2(fc)	pval	regulation
FAT4	-2.33	0.00	down
AP000654	-3.22	0.00	down
KLRC4	-2.43	0.00	down
PCDHGB7	-3.18	0.00	down
LINC00299	-1.77	0.00	down
GNLY	-1.88	0.00	down
AC034102	-2.46	0.00	down
KLRC3	-2.62	0.00	down
OR2L6P	-3.25	0.00	down
AC243829	-2.44	0.00	down
KLRC2	-3.05	0.00	down
ENPP5	-1.79	0.00	down
OR2L2	-3.97	0.00	down
AC006033	-1.72	0.00	down
BNC2	-2.16	0.00	down
GPR141BP	-2.11	0.00	down
AL365475	-1.92	0.00	down
LINC00893	-1.72	0.00	down
MIR181A2HG	-2.58	0.00	down



(a)



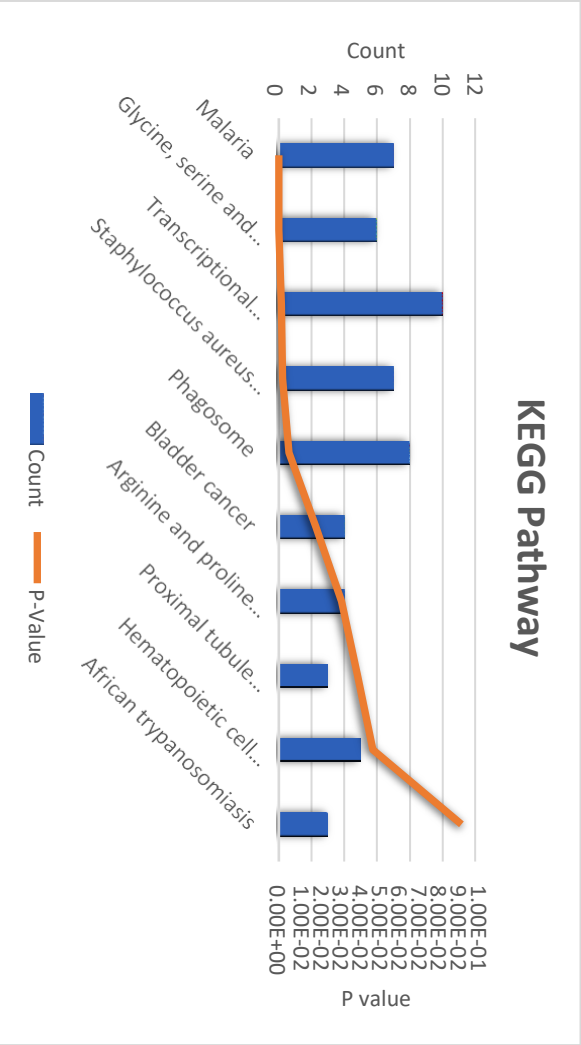
(b)



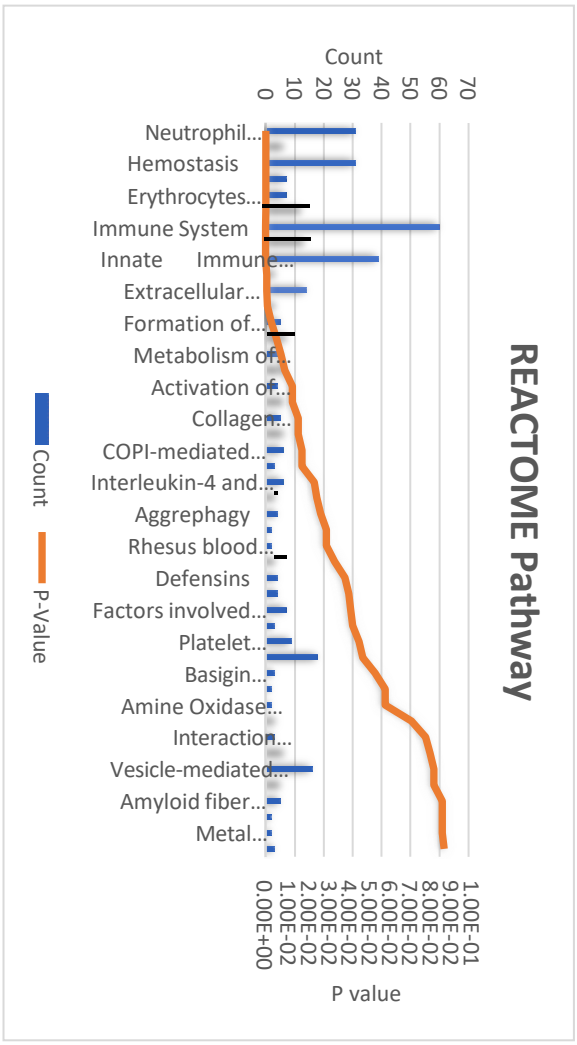
(c)

Fig 5.1: Enrichment analysis of up-regulated genes performed by evaluating the Gene ontology terms.

(a) Biological Process; (b) Molecular Function; (c) Cellular Component

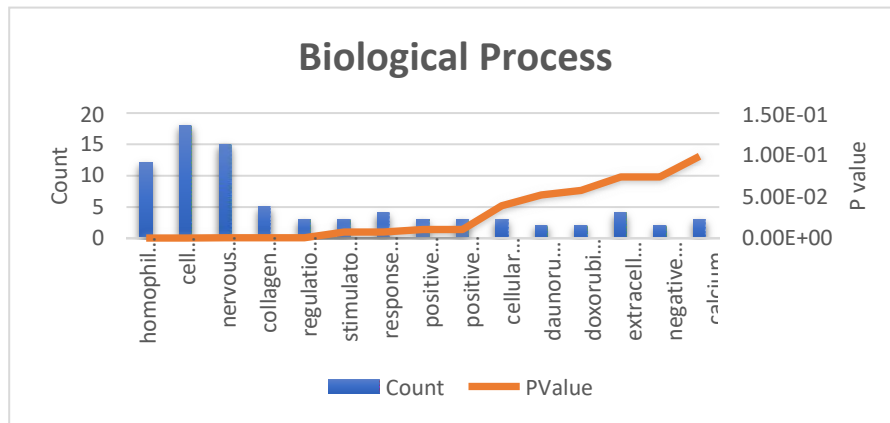


(a)

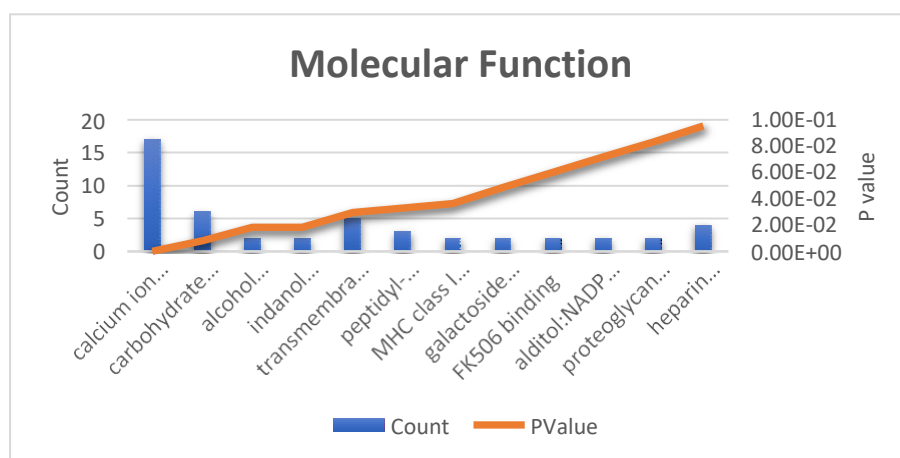


(b)

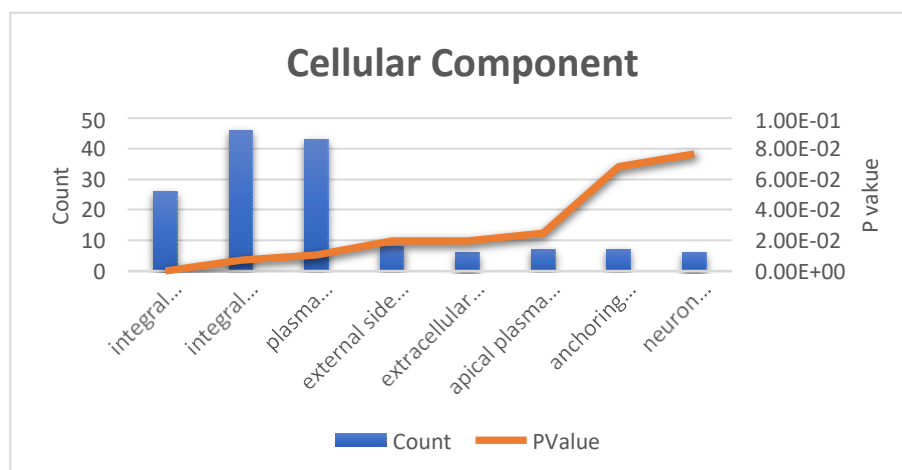
Fig 5.2: Enrichment analysis of up-regulated genes performed by evaluating the pathway associations.
(a) KEGG pathway; (b) REACTOME pathway



(a)

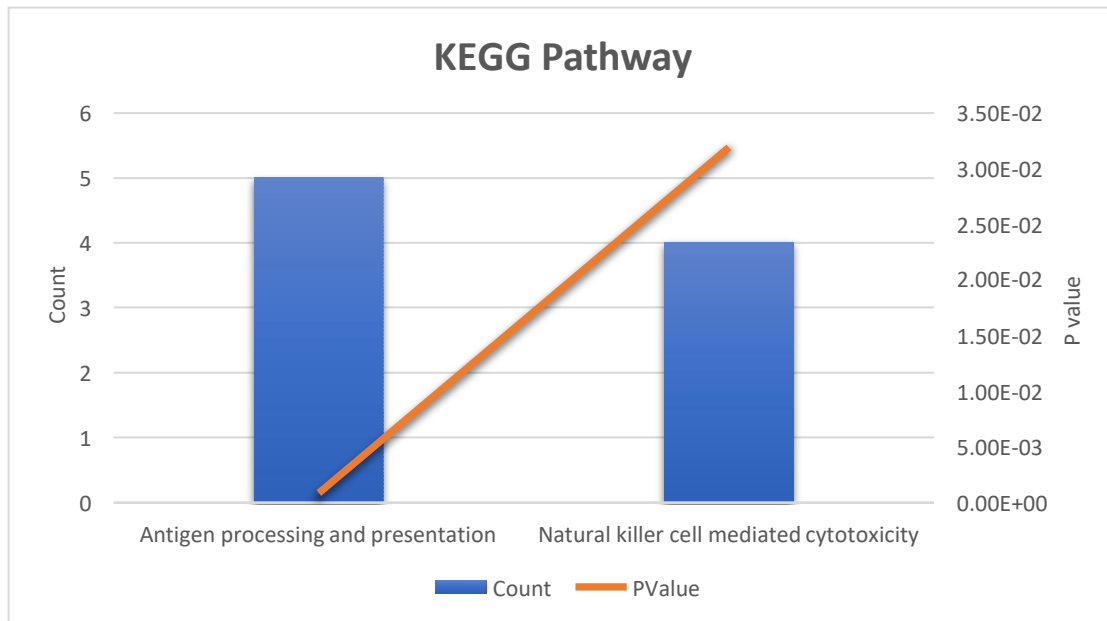


(b)

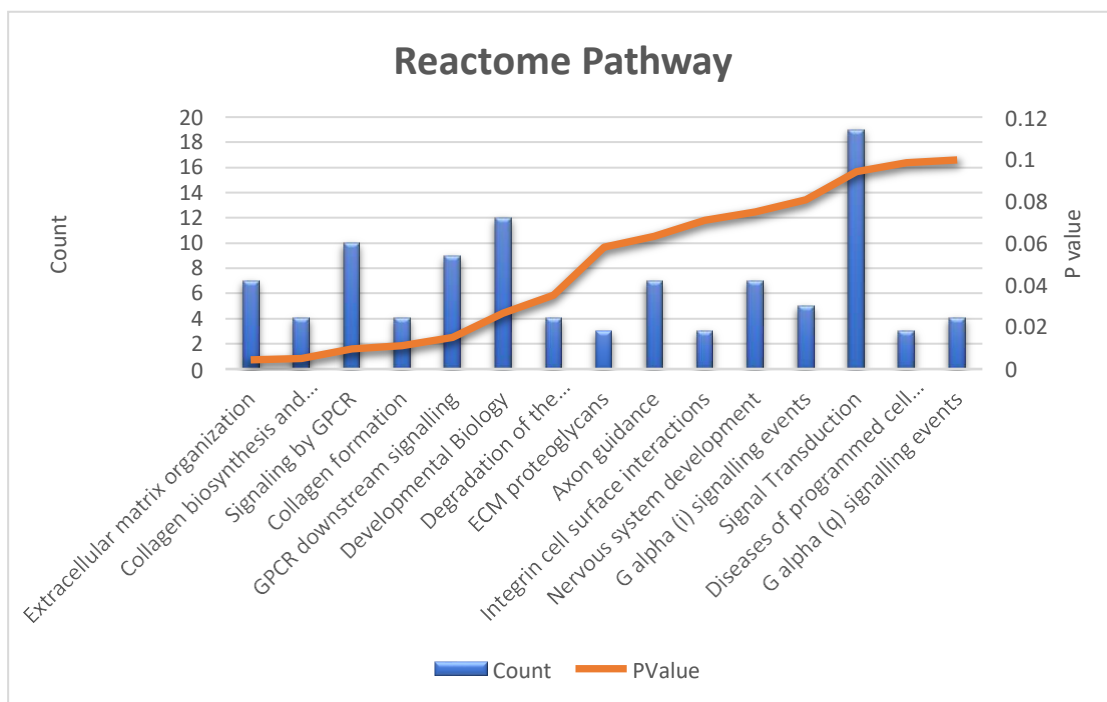


(c)

Fig 5.3: Enrichment analysis of down-regulated genes performed by evaluating the Gene ontology terms.
(a) Biological Process; (b) Molecular Function; (c) Cellular Component



(a)



(b)

Fig 5.4: Enrichment analysis of down-regulated genes performed by evaluating the pathway associations.

(a) KEGG pathway; (b) REACTOME pathway

5.3 PPI NETWORK CONSTRUCTION:

The up-regulated and down-regulated genes were uploaded in the STRING database separately. The PPI network for up-regulated genes was constructed with 251 nodes and 647 edges on an average node degree 5.16, average local clustering co-efficient 0.416 and PPI enrichment p-value $<1.0e-16$ (Fig 5.5a). The PPI network for down-regulated genes was constructed with 232 nodes and 199 edges on an average node degree 1.72, average local clustering coefficient 0.387 and PPI enrichment p-value $6.23e-08$ (Fig 5.5b).

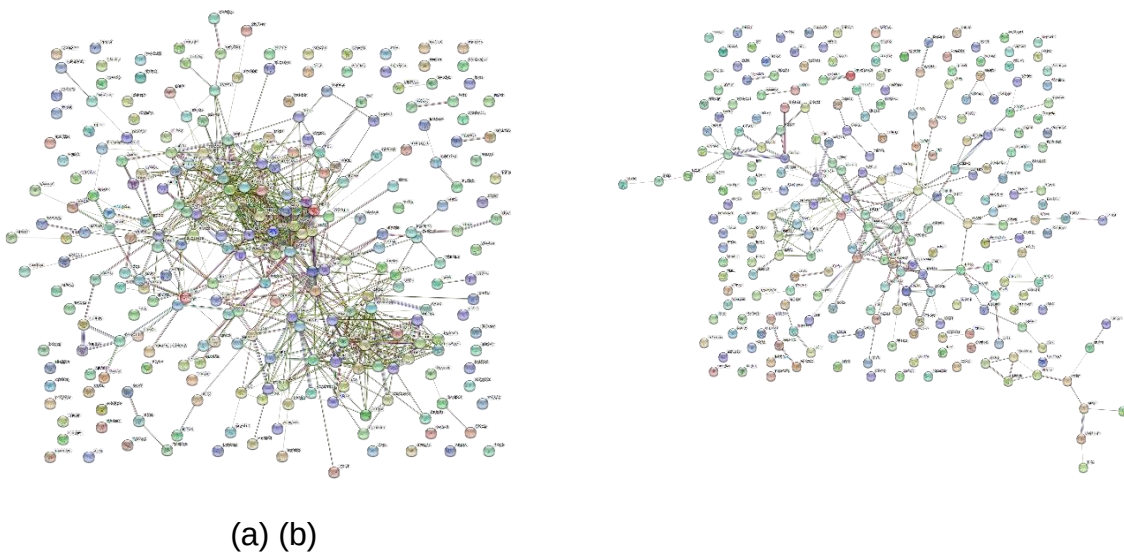
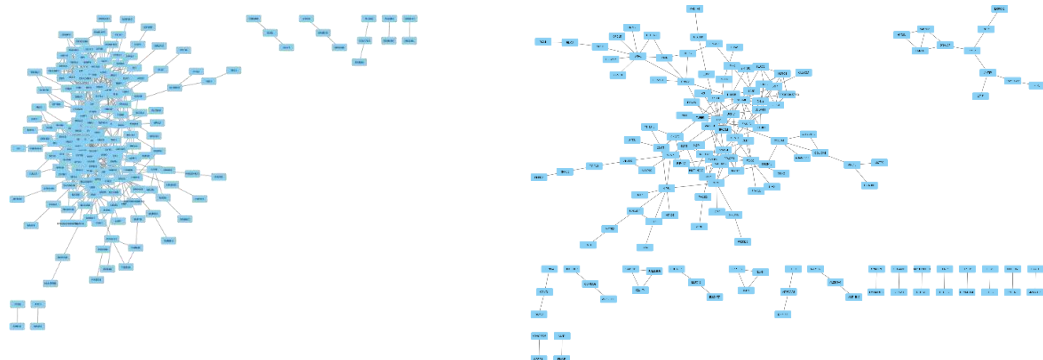


Fig 5.5: PPI network constructed using STRING database (a) Up-regulated genes network; (b) Down-regulated genes network

5.4 SCREENING OF PPI FOR MODULES AND HUB GENES ANALYSIS:

The networks obtained from STRING were exported to cytoscape to visualize the interaction among the query proteins and to identify the hub genes (Fig 5.6). MCODE plugin was used to identify the clusters containing the closely inter-linked nodes in the network. A cluster containing 13 nodes and 75 edges with a highest cluster score of 12.500 was considered among the 7 different clusters found from the MCODE results in the up-regulated genes network (Fig 5.7a). For down-regulated genes network, a cluster containing 4 nodes and 5 edges with a highest cluster score of 3.333 was considered among the 6 different clusters found from

the MCODE results (Fig 5.7b). Another plugin of cytoscape called CytoHubba, that uses 12 different algorithms and provides a user-friendly interface to analyse the topology of PPI networks was used to predict the hub genes. Cytohubba was performed on the resulting clusters of MCODE and the entire PPI network separately. The resulting genes from each algorithm were first compared with those from the other 12 algorithms and a list of 8 common genes were obtained from the up-regulated genes network (Fig 5.8) and 4 common genes from the down-regulated genes network (Fig 5.9), which were again compared with the results of MCODE. The same method was performed on the entire PPI network and the results were compared with the previous CytoHubba results (Fig 5.10) (Fig 5.11). The CytoHubba results revealed that there are 20 hub genes in total and out of which 10 genes are up-regulated (HBB, ALAS2, GATA1, AHSP, HBG1, HBG2, HBD, KLF1, SLC4A1, EPB42) and 10 genes are down-regulated (ZMYND10, DNAJC7, HYDIN, LRRC6, FN1, NCAM1, FASLG, CTCF, SMAD4, STAT1).

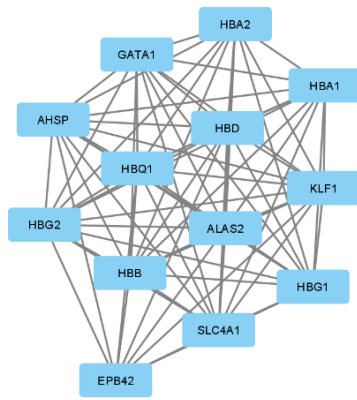


(a) Up-regulated network

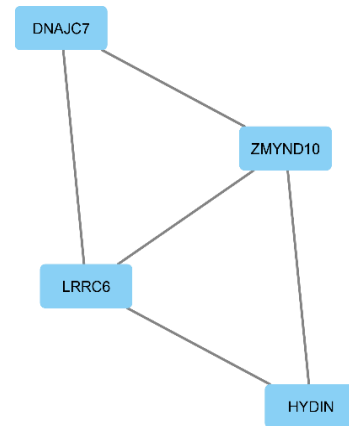
(b) Down-regulated network

Fig 5.6: Visualization of PPI network using Cytoscape

Note: The nodes represent the proteins and the edges represent the interactions of the proteins.



(a) MCODE Cluster for
Up-regulated genes



(b) MCODE Cluster for
Down-regulated genes

Fig 5.7: Module analysis using MCODE

Note: The selection standards of MCODE are Network with highest MCODE Score; Degree cut-off = 2; Node score cut-off = 0.2; K-score = 2; Max depth = 100.

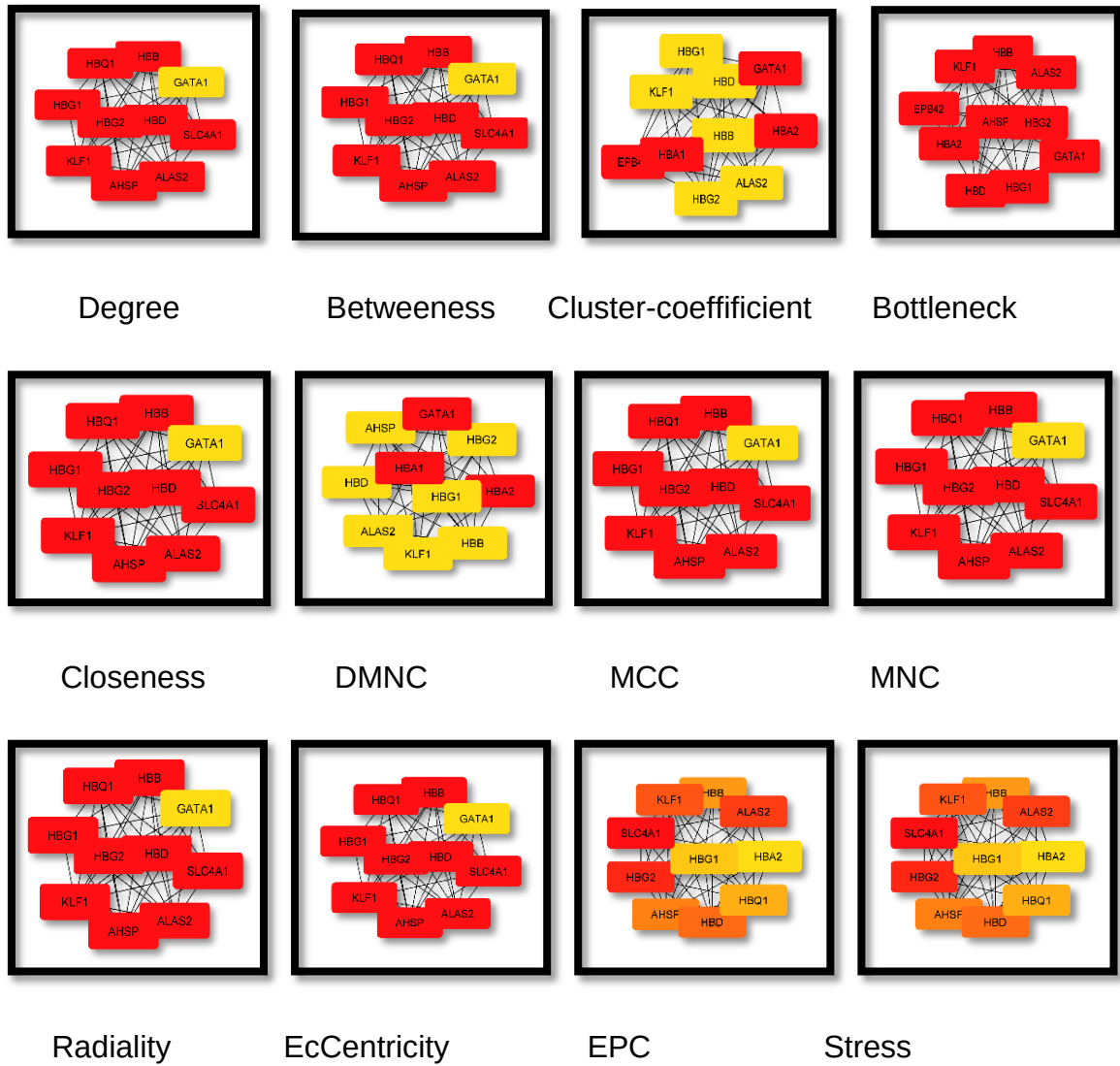


Fig 5.8: Common hub genes in up-regulated network from CytoHubba on MCODE

Note: Colours ranging from red to yellow are used to indicate the rank (1-10) of the connection degree.



Fig 5.9: Common hub genes in down-regulated network from CytoHubba on MCODE

Note: Colours ranging from red to yellow are used to indicate the rank (1-10) of the connection degree.

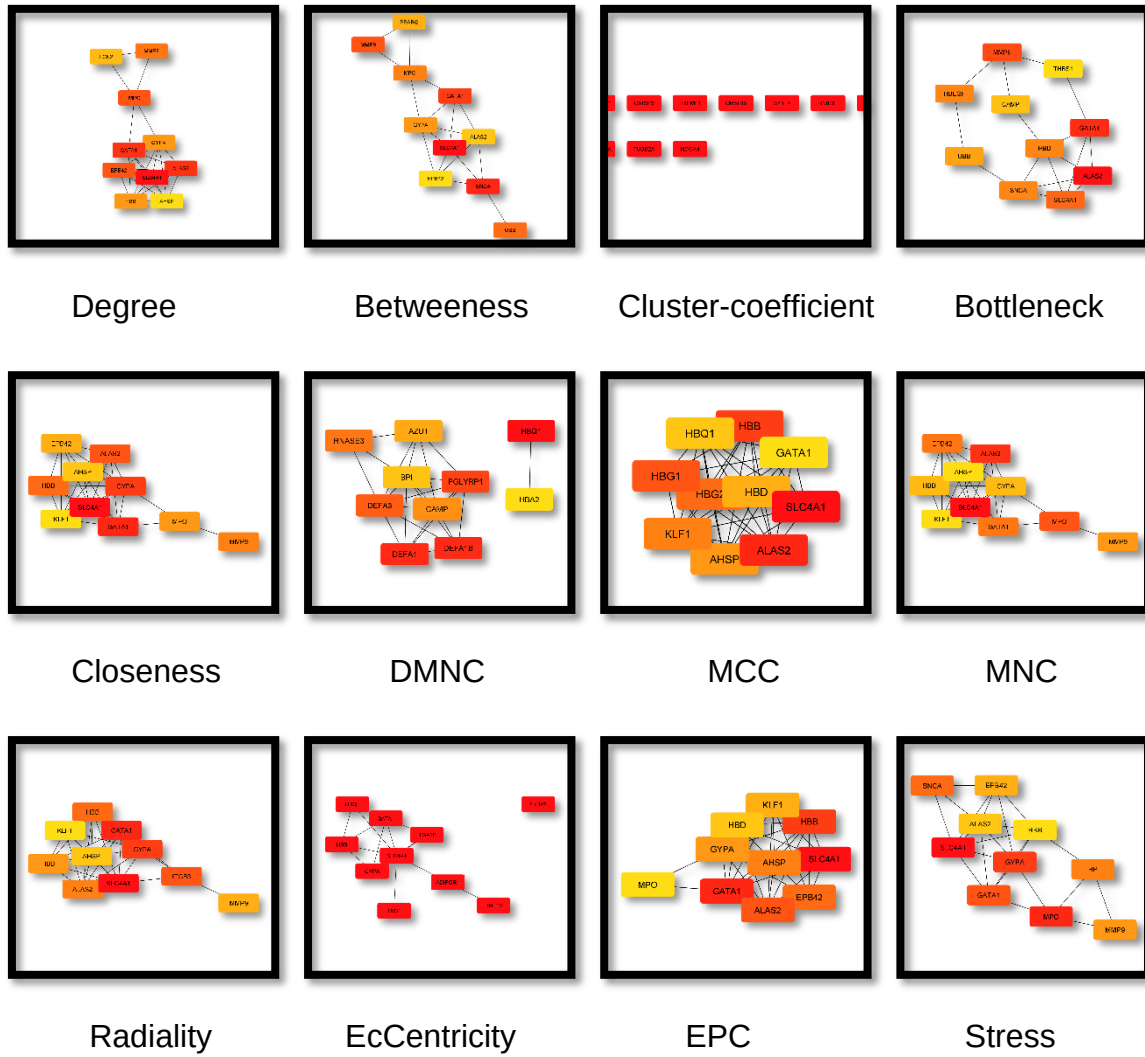


Fig 5.10: Common hub genes in up-regulated network from CytoHubba on entire network

Note: Colours ranging from red to yellow are used to indicate the rank (1-10) of the connection degree.

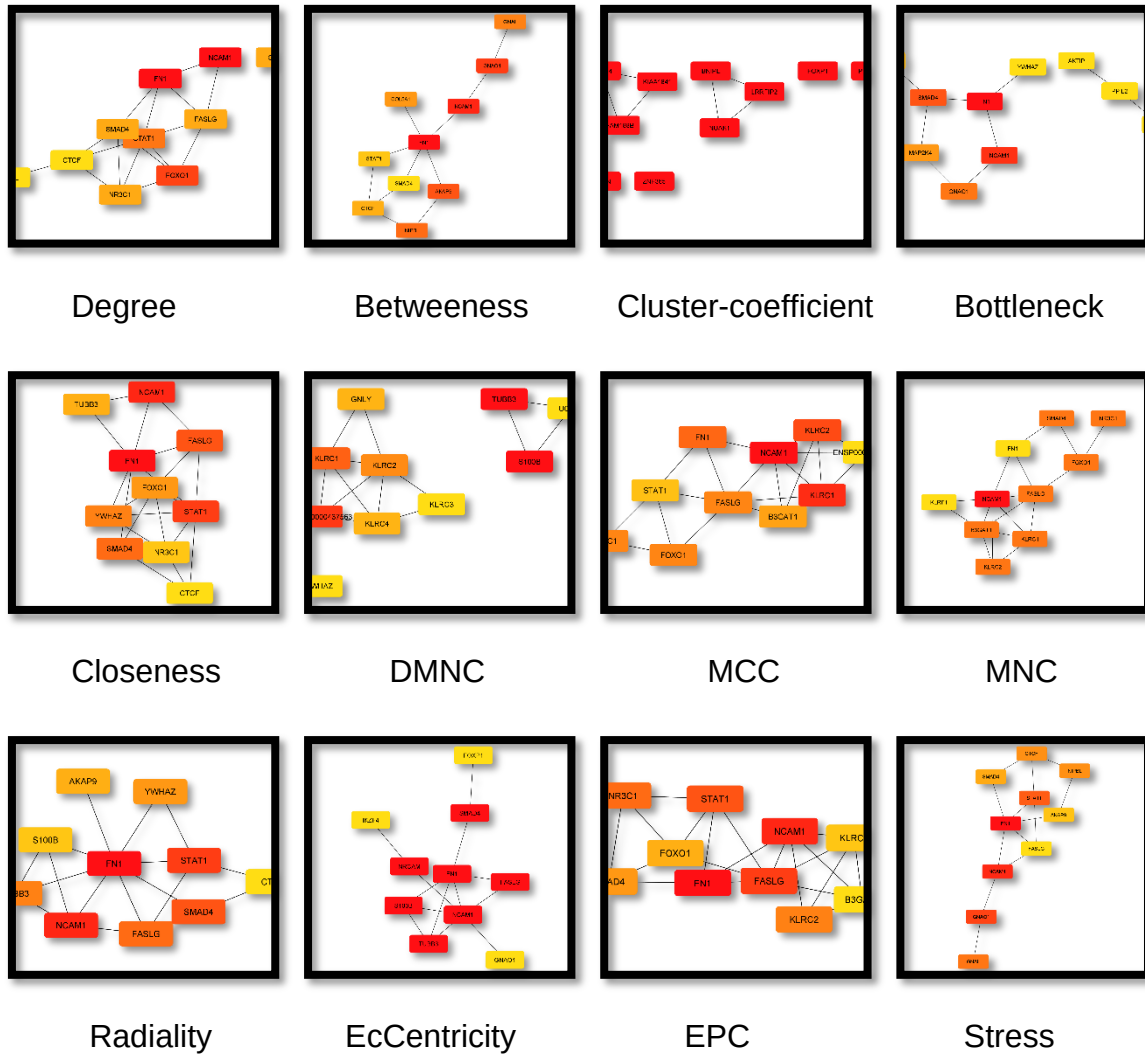


Fig 5.11: Common hub genes in down-regulated network from CytoHubba on entire network

Note: Colours ranging from red to yellow are used to indicate the rank (1-10) of the connection degree.

5.5 GENE EXPRESSION VALIDATION AND DISEASE ASSOCIATION ANALYSIS:

The results of DisGeNET indicated the fact that the predicted hub genes are mainly associated with Immune systems disorders, arthritis, Skin & Connective Tissue diseases and Musculoskeletal diseases, while STAT1, FASLG, FN1, NCAM1 and GATA1 are directly associated with Rheumatoid arthritis (Table 5.2). The descriptions and functions of the predicted hub genes are provided in Table 5.3 and their association with other diseases are provided in Table 5.4.

Table 5.2: Gene-disease association

Gene	Disease	Type	Disease Class	Semantic Type	N. genes d	N. SNPs d	.l
STAT1	Rheumatoid Arthritis	disease	Skin and Connective Tissue Diseases; Musculoskeletal Diseases; Immune System Diseases	Disease or Syndrome	2723	2387	0.4
FASLG	Rheumatoid Arthritis	disease	Skin and Connective Tissue Diseases; Musculoskeletal Diseases; Immune System Diseases	Disease or Syndrome	2723	2387	0.34
FN1	Rheumatoid Arthritis	disease	Skin and Connective Tissue Diseases; Musculoskeletal Diseases; Immune System Diseases	Disease or Syndrome	2723	2387	0.1
NCAM1	Rheumatoid Arthritis	disease	Skin and Connective Tissue Diseases; Musculoskeletal Diseases;	Disease or Syndrome	2723	2387	0.03

			tal Diseases; Immune System Diseases				
GATA1	Rheumatoid Arthritis	diseas e	Skin and Connective Tissue Diseases; Musculoskele tal Diseases; Immune System Diseases	Disease or Syndrome	2723	2387	0.01

Table 5.3:

(a) Functions of the hub genes found from up-regulated network

Symbol	Description	Functions
HBG1	hemoglobin subunit gamma 1 [Source:HGNC Symbol;Acc:HGNC:4831]	antioxidant activity, coagulation, cytosol, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, acting on peroxide as acceptor, reactive oxygen species metabolic process, tetrapyrrole binding
HBG2	hemoglobin subunit gamma 2 [Source:HGNC Symbol;Acc:HGNC:4832]	antioxidant activity, coagulation, cytosol, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, acting on peroxide as acceptor, reactive oxygen species metabolic process, tetrapyrrole binding
AHSP	alpha hemoglobin stabilizing protein [Source:HGNC Symbol;Acc:HGNC: 18075]	cytosol, erythrocyte homeostasis, homeostasis of number of cells, myeloid cell differentiation, myeloid cell homeostasis
HBD	hemoglobin subunit delta [Source:HGNC Symbol;Acc:HGNC:4829]	antioxidant activity, coagulation, cytosol, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, acting on peroxide as acceptor, reactive oxygen species metabolic process, tetrapyrrole binding

ALAS2	5'-aminolevulinate synthase 2 [Source:HGNC Symbol;Acc:HGNC:397]	erythrocyte differentiation, erythrocyte homeostasis, homeostasis of number of cells, myeloid cell development, myeloid cell differentiation, myeloid cell homeostasis
HBB	hemoglobin subunit beta [Source:HGNC Symbol;Acc:HGNC:4827]	antioxidant activity, coagulation, cytosol, gas transport, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, acting on peroxide as acceptor, reactive oxygen species metabolic process, tetrapyrrole binding
SLC4A1	solute carrier family 4 member 1 (Diego blood group) [Source:HGNC Symbol;Acc:HGNC:11027]	chloride/bicarbonate exchanger involved in carbon dioxide transport from tissues to lungs, hemoglobin binding, erythrocyte development
EPB42	erythrocyte membrane protein band 4.2 [Source:HGNC Symbol;Acc: HGNC:3381]	ATP-binding protein which may regulate the association of protein 3 with ankyrin; has a role in erythrocyte shape and mechanical property regulation; hemoglobin metabolic process
KLF1	Kruppel like factor 1 [Source:HGNC Symbol;Acc:HGNC:6345]	erythrocyte homeostasis, homeostasis of number of cells, myeloid cell differentiation, myeloid cell homeostasis
GATA1	GATA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:4170]	coagulation, erythrocyte differentiation, erythrocyte homeostasis, hemostasis, homeostasis of number of cells, myeloid cell development, myeloid cell differentiation, myeloid cell homeostasis

(b) Functions of the hub genes found from down-regulated network

Symbol	Description	Functions
HYDIN	HYDIN axonemal central pair apparatus protein [Source:HGNC Symbol; Acc:HGNC:19368]	cilium movement
DNAJC7	DnaJ heat shock protein family (Hsp40) member C7 [Source:HGNC Symbol;Acc:HGNC:12392]	heat shock protein binding activity, chaperone cofactor-dependent protein refolding
CTCF	CCCTC-binding factor [Source:HGNC Symbol;Acc:HGNC:13723]	transcription factor activity, sequence-specific DNA binding, protein binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, transcription regulatory region sequence-specific DNA binding

ZMYND10	zinc finger MYND-type containing 10 [Source:HGNC Symbol;Acc: HGNC:19412]	axonemal dynein complex assembly, axoneme assembly
FASLG	Fas ligand [Source:HGNC Symbol;Acc:HGNC:11936]	tumor necrosis factor receptor binding, protein binding, negative regulation of transcription from RNA polymerase II promoter, cellular response to interferon-gamma
LRRC6	leucine rich repeat containing 6 [Source:HGNC Symbol;Acc:HGNC: 16725]	axonemal dynein complex assembly, axoneme assembly
FN1	fibronectin 1 [Source:HGNC Symbol;Acc:HGNC:3778]	cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defense, and metastasis
STAT1	signal transducer and activator of transcription 1 [Source:HGNC Symbol; Acc:HGNC:11362]	cellular response to interleukin-6, response to interleukin-6
NCAM1	neural cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:7656]	involved in cell-to-cell interactions as well as cell-matrix interactions during development and differentiation; plays a role in the development of the nervous system by regulating neurogenesis, neurite outgrowth, and cell migration; involved in the expansion of T lymphocytes, B lymphocytes and natural killer (NK) cells which play an important role in immune surveillance; plays a role in signal transduction by interacting with fibroblast growth factor receptors, N-cadherin and other components of the extracellular matrix and by triggering signalling cascades involving FYN-focal adhesion kinase (FAK), mitogen-activated protein kinase (MAPK), and phosphatidylinositol 3-kinase (PI3K)
SMAD4	SMAD family member 4 [Source:HGNC Symbol;Acc:HGNC:6770]	cellular response to interleukin-6, response to interleukin-6

Table 5.4: Genes association with other diseases

Gene	Associated Diseases
HBG1	Fetal hemoglobin quantitative trait locus 1
HBG2	Cyanosis transient neonatal (TNCY); Fetal hemoglobin quantitative trait locus 1
AHSP	beta-thalassemia
HBD	beta-thalassemia
ALAS2	Anemia, sideroblastic, 1 (SIDBA1); Erythropoietic protoporphyria, X-linked dominant (XLDPT)
HBB	Hereditary hemolytic anemia; Congenital dyserythropoietic anemia; Beta thalassemia; Sickle cell disease; Methemoglobinemia, beta-globin type
SLC4A1	SLC4A1-associated distal renal tubular acidosis, Hereditary spherocytosis, Hereditary hemolytic anemia
EPB42	Hereditary hemolytic anemia, Spherocytosis 5 (SPH5)
KLF1	Anemia, congenital dyserythropoietic, 4 (CDAN4); Hereditary hemolytic anemia; Congenital dyserythropoietic anemia
GATA1	X-linked dyserythropoietic anemia and thrombocytopenia (XDAT); Thrombocytopenia with beta-thalassemia, X-linked (XLTT); Anemia without thrombocytopenia, X-linked (XLAWT); Hemolytic anemia due to elevated adenosine deaminase (HAEADA); Diamond-Blackfan anemia
HYDIN	Kartagener syndrome; Primary ciliary dyskinesia; Polynesian Bronchiectasis; Precursor T-Cell Lymphoblastic Leukemia-Lymphoma; Miscarriage; Spontaneous abortion; Early Pregnancy Loss
DNAJC7	Breast carcinoma

CTCF	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma; Intellectual developmental disorder, autosomal dominant 21 (MRD21)
ZMYND10	Kartagener syndrome; Primary ciliary dyskinesia; Polynesian Bronchiectasis
FASLG	Autoimmune lymphoproliferative syndrome 1A (ALPS1A); systemic lupus erythematosus (SLE)
LRRC6	Ciliary dyskinesia, primary, 19 (CILD19)
FN1	Glomerulopathy with fibronectin deposits 2 (GFND2); Spondylometaphyseal dysplasia; corner fracture type (SMDCF); Miscarriage; Spontaneous abortion; Early Pregnancy Loss; Abortion, Tubal
STAT1	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma; Immunodeficiency 31B (IMD31B); Immunodeficiency 31A (IMD31A); Immunodeficiency 31C (IMD31C)
NCAM1	Miscarriage; Spontaneous abortion; Early Pregnancy Loss; Abortion, Tubal
SMAD4	Hereditary hemorrhagic telangiectasia; Juvenile polyposis syndrome; Myhre syndrome; Cholangiocarcinoma

5.6 IDENTIFICATION OF miRNAs TARGETING THE HUB GENES:

The miRNAs and transcription factors (TF) targeting hub genes were predicted using miRNet 2.0 platform. A query list of 20 genes were submitted to the database and 2 genes (AHSP, ZMYND10) were not mapped to the interaction database. In the resulting network, there are 18 genes, 70 transcription factors and 402 miRNAs with 660 edges found (Fig 5.12). Using a degree filter (network tool), it was possible to eliminate the less significant nodes and find the nodes with higher degrees that serve as significant hubs in the network. The network after applying degree filter revealed 18 genes, 103 miRNAs and 10 transcription factors with 301 edges (Fig 5.13). Additionally, the network demonstrated that the hub genes GATA1 and KLF1 are both transcription factors and genes. Out of the 402 miRNAs that were produced, hsa-mir-27a-3p targeted 10 hub genes and out of the

70 transcription factors that were produced, GATA1, NFKB1, and RELA targeted 6 hub genes.

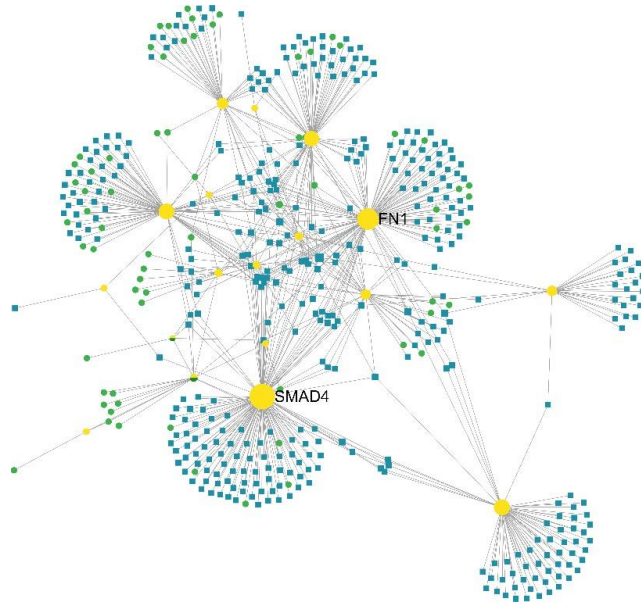


Fig 5.12: Network of regulatory interactions between miRNAs, transcription factors, and hub genes

Note: Yellow circles represent the Hub genes, Green circles represent the transcription factors and Blue squares represent the miRNAs in the network.

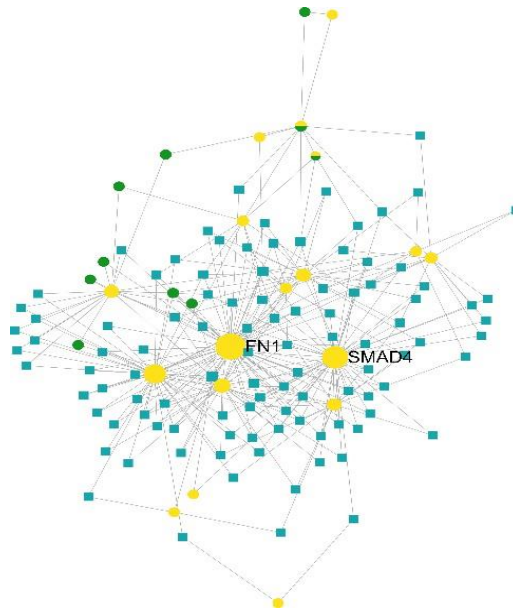


Fig 5.13: Filtered network of miRNAs and transcription factors targeting hub genes

Note: Yellow circles represent the Hub genes, Green circles represent the transcription factors and Blue squares represent the miRNAs in the network.

5.6 DISCUSSION:

In this study, the RNA sequencing dataset GSE193193 containing 5 RA samples and 10 control samples was obtained from the GEO database and analysed to screen the Differentially Expressed Genes (DEGs) in Rheumatoid arthritis. From a total of 36,107 DEGs found in the dataset, we examined that 3672 DEGs were significant based on the adjusted p value (<0.05) and 283 genes were significantly up-regulated and 360 genes were significantly down-regulated based on the log2fc values (>1.5 and <-1.5) (Table 5.1).

The Gene ontology enrichment analysis performed using DAVID tool revealed that the up-regulated genes involved in Biological Process were mostly enriched in innate immune response, negative regulation of apoptotic process and cell adhesion (Fig 5.1), while the down-regulated genes were again enriched in cell adhesion, nervous system development and homophilic cell adhesion via plasma membrane adhesion molecules (Fig 5.3). Another GO term analysis, Molecular Function revealed that the up-regulated genes were mostly enriched in identical

protein binding and zinc ion binding (Fig 5.1), while the down-regulated genes were significantly enriched in calcium ion binding (Fig 5.3). By analysing the Cellular Component term, we found that the up-regulated genes were mostly linked with plasma membrane, cytosol, and extracellular exosome (Fig 5.1), while the down-regulated genes were again linked with plasma membrane and integral component of membrane (Fig 5.3). The KEGG and REACTOME pathway analysis revealed that the up-regulated genes were mostly involved in transcriptional mis-regulation in cancer, phagosome, immune system and innate immune system (Fig 5.2), while the down-regulated genes were mostly involved in antigen processing & presentation, natural killer cell mediated cytotoxicity, signal transduction and developmental biology (Fig 5.4). By performing enrichment analysis we have identified the biological processes and pathways that are significantly related to the genes under study, thereby providing a cue to their involvement in disease.

To further screen the DEGs, the PPI networks were constructed for the up-regulated and down-regulated genes separately using STRING database. The PPI network for up-regulated genes was constructed with 251 nodes and 647 edges (Fig 5.5), while PPI network for the down-regulated genes was constructed with 232 nodes and 199 edges (Fig 5.5) and these networks were exported to Cytoscape software and visualized for identifying the molecular interaction (Fig 5.6). The PPI networks were further screened using MCODE plugin to identify the most significant cluster modules from the entire network, which is distinguished by its highest score and the highly interconnected regions (Fig 5.7). The significant clusters were considered also based on some selection standards such as degree cut-off = 2, node score cut-off = 2, K-score = 2 and max depth = 100. Screening the gene clusters could aid in the prediction of core genes and their interactions, as well as how they are linked with the pathogenesis and progression of Rheumatoid arthritis.

Further analysis of the clusters and the entire PPI networks using CytoHubba revealed that there are 20 hub genes in total and out of which 10 genes (HBB, ALAS2, GATA1, AHSP, HBG1, HBD, KLF1, HBG2, SLC4A1, EPB42) were up-regulated (Fig 5.8) (Fig 5.10) and 10 genes (ZMYND10, DNAJC7, HYDIN, LRRC6, FN1, NCAM1, FASLG, CTCF, SMAD4, STAT1) were down-regulated (Fig 5.9) (Fig 5.11).

Hemoglobin subunit beta (HBB) is a gene that provides instructions for making a protein called hemoglobin. HBB is involved in oxygen transport from the lung to the various peripheral tissues (UniProt). It is reported that the normal adult hemoglobin tetramer consists of two alpha chains and two beta chains where, mutant beta globin causes sickle cell anemia; absence of beta chain causes beta-zero-thalassemia; and reduced amounts of detectable beta globin causes beta-plus-thalassemia (NCBI). HBB is generally involved in antioxidant activity, coagulation, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, reactive oxygen species metabolic process, etc (GeneMania). According to DisGeNET database, HBB is also involved in musculoskeletal diseases such as Knee joint valgus deformity. From this we can understand that HBB is indirectly associated in the onset of Rheumatoid arthritis.

ALAS2 gene provides instructions for making an enzyme called 5'-aminolevulinate synthase 2 or erythroid ALA-synthase. This version of the enzyme is found only in developing red blood cells called erythroblasts (MedlinePlus Genetics). ALAS2 catalyzes the pyridoxal 5'-phosphate (PLP) – dependent condensation of succinyl-CoA and glycine to form aminolevulinic acid (ALA), with CoA and CO₂ as by-products. It contributes significantly to heme formation during erythropoiesis (UniProt). ALAS2 generally involves the functions of erythrocyte differentiation, erythrocyte homeostasis, homeostasis of number of cells, myeloid cell development, myeloid cell differentiation, myeloid cell homeostasis (GeneMania).

The GATA binding protein 1 (GATA1) gene is a transcriptional activator or repressor which serves as a general switch factor for erythroid development. It binds to DNA sites with the consensus sequence 5'-[AT]GATA[AG]-3' within regulatory regions of globin genes and of other genes expressed in erythroid cells. It activates the transcription of genes involved in erythroid differentiation of K562 erythroleukemia cells, including HBB, HBG1/2 and ALAS2 (UniProt). GATA1 is generally involved in coagulation, erythrocyte differentiation, erythrocyte homeostasis, homeostasis of number of cells, myeloid cell development, myeloid cell differentiation and myeloid cell homeostasis (GeneMania). According to DisGeNET, GATA1 gene is involved in musculoskeletal diseases such as Foot deformities and Hand deformities. It is also involved in Immunological deficiency

syndromes and Skin and Connective Tissue Diseases such as Rheumatoid arthritis.

The Alpha-hemoglobin-stabilizing protein (AHSP) acts as a chaperone to prevent the harmful aggregation of alpha-hemoglobin during normal erythroid cell development. Specifically protects free alpha-hemoglobin from precipitation. It is predicted to modulate pathological states of alpha-hemoglobin excess such as beta-thalassemia (UniProt). It is generally involved in cystol, erythrocyte homeostasis, homeostasis of number of cells, myeloid cell differentiation and myeloid cell homeostasis (GeneMania).

Hemoglobin subunit gamma-1 (HBG1) and Hemoglobin subunit gamma-2 are normally expressed in the fetal liver, spleen and bone marrow. It is reported that the gamma globin genes (HBG1 and HBG2) together with two alpha chains constitute fetal hemoglobin (HbF) which is normally replaced by adult hemoglobin (HbA) at birth. In some beta-thalassemias and related conditions, gamma chain production continues into adulthood (NCBI). HBG1 and HBG2 are generally involved in antioxidant activity, coagulation, cystol, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, reactive species metabolic process and tetrapyrrole binding (GeneMania).

The Hemoglobin subunit delta (HBD) is involved in oxygen transport from the lung to the various peripheral tissues (UniProt). Mutations in the delta-globin gene are associated with beta-thalassemia (NCBI). It is generally involved in antioxidant activity, coagulation, cystol, hemostasis, hydrogen peroxide metabolic process, molecular carrier activity, oxidoreductase activity, acting on peroxide as acceptor, reactive oxygen species metabolic process and tetrapyrrole binding (GeneMania).

The Krueppel-like factor 1 (KLF1) is a transcription regulator of erythrocyte development that probably serves as a general switch factor during erythropoiesis. It is a dual regulator of fetal-to-adult globin switching; binds to the CACCC box in the beta-globin gene promoter and acts as a preferential activator of this gene. Furthermore, it binds to the BCL11A promoter and activates expression of BCL11A, which in turn represses the HBG1 and HBG2 genes. This dual activity ensures that, in most adults, fetal hemoglobin levels are low. It is able to activate CD44 and AQP1 promoters. KLF1, when sumoylated, acts as a transcriptional

repressor by promoting interaction with CDH2/MI2beta and also represses megakaryocytic differentiation (UniProt). It is generally involved in erythrocyte homeostasis, homeostasis of number of cells, myeloid cell differentiation and myeloid cell homeostasis (GeneMania).

Solute Carrier family 4 member 1 (SLC4A1) gene provides instructions for making a protein known as anion exchanger 1 (AE1). This protein transports negatively charged atoms (anions) across cell membranes. Specifically AE1 exchanges negatively charged atoms of chlorine (chloride ions) for negatively charged bicarbonate molecules (bicarbonate ions). Based on this function, AE1 is known as a chloride / bicarbonate exchanger ($\text{Cl}^-/\text{HCO}_3^-$ exchanger). The main function of this exchanger is to maintain the correct acid levels (pH) in the body. There are two versions of the AE1 protein that differ in size. The shorter version is found in specialized kidney cells, called alpha-intercalated cells, that line structures in the kidney called renal tubules (MedlinePlus Genetics). Mutations in this gene leads to two types of diseases: destabilization of red cell membrane leading to hereditary spherocytosis, and defective kidney acid secretion leading to distal renal tubular acidosis. Other mutations that do not give rise to disease result in novel blood group antigens, which form the Diego blood group system. Southeast Asian ovalocytosis (SAO, Melanesian ovalocytosis) results from the heterozygous presence of a deletion in the encoded protein and is common in areas where plasmodium falciparum malaria is endemic. One null mutation in this gene is known, resulting in very severe anaemia and nephrocalcinosis (NCBI).

Erythrocyte membrane protein band 4.2 (EPB42) is an ATP-binding protein which may regulate the association of protein 3 with ankyrin. It probably has a role in erythrocyte shape and mechanical property regulation. Mutations in the EPB42 gene are associated with recessive spherocytic elliptocytosis and recessively transmitted hereditary anaemia. Alternatively spliced transcript variants encoding different isoforms have been found for this gene (NCBI).

Zinc finger MYND-type containing 10 (ZMYND10) gene encodes a protein containing a MYND-type zinc finger domain that likely functions in assembly of the dynein motor. Mutations in this gene can cause primary ciliary dyskinesia. This gene is also considered a tumour suppressor gene and is often mutated, deleted,

or hypermethylated and silenced in cancer cells. Alternatively splicing results in multiple transcript variants (NCBI).

DnaJ heat shock protein family (Hsp40) membrane C7 (DNAJC7) gene encodes a member of the DNAJ heat shock protein 40 family of proteins that is characterized by two N-terminal tetratricopeptide repeat domains and a C-terminal DNAJ domain. This protein binds the chaperone proteins heat shock proteins 70 and 90 in an ATP-dependent manner and may function as a co-chaperone. It is reported that pseudogenes of this gene are found on chromosomes 1 & 6 and alternate splicing results in multiple transcript variants (NCBI).

Hydrocephalus-inducing protein homolog (HYDIN) is mainly required for ciliary motility (UniProt). Mutations in this gene causes autosomal recessive primary ciliary dyskinesia-5, a disorder characterized by the accumulation of cerebrospinal fluid within the ventricles of the brain. A duplicate copy of this gene has been found in humans on chromosome 1 (NCBI).

Leucine rich repeat containing 6 (LRRC6) is involved in dynein arm assembly and is important for expression and transporting outer dynein arm (ODA) proteins from the cytoplasm to the cilia. It acts as a crucial component in the formation and motility of spermatozoal flagella (UniProt). Defects in this gene are a cause of primary ciliary dyskinesia-19 (CILD19). Alternative splicing of this gene results in multiple transcripts variants. Related pseudogenes have been identified on chromosomes 4, 11 and 22 (NCBI).

Fibronectin 1 (FN1) is involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape. Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. It is also involved in osteoblast compaction through the fibronectin fibrillogenesis cell-mediated matrix assembly process, essential for osteoblast mineralization; participates in the regulation of type 1 collagen deposition by osteoblasts; acts as a ligand for the LILRB4 receptor, inhibiting FCGR1A/CD64-mediated monocyte activation (UniProt). According to DisGeNET, FN1 gene is involved in Skin & Connective Tissue diseases such as Rheumatoid arthritis and Musculoskeletal diseases such as Degenerative polyarthritis and knee joint valgus deformity. FN1

gene is also observed in pathological conditions such as Hip joint varus deformity and Edema of foot.

Neural Cell Adhesion Molecule 1 (NCAM1) gene encodes a cell adhesion protein which is a member of the immunoglobulin superfamily. The encoded protein is involved in cell-to-cell interactions as well as cell-matrix interactions during development and differentiation. The encoded protein plays a role in the development of the nervous system by regulating neurogenesis, neurite outgrowth, and cell migration. This protein is also involved in the expansion of T lymphocytes, B lymphocytes and natural killer (NK) cells which play an important role in immune surveillance. This protein plays a role in signal transduction by interacting with fibroblast growth factor receptors, N-cadherin and other components of the extracellular matrix and by triggering signalling cascades involving FYN-focal adhesion kinase (FAK), mitogen-activated protein kinase (MAPK), and phosphatidylinositol 3-kinase (PI3K) (NCBI). NCAM1 acts as a receptor for rabies virus and zika virus (UniProt). According to DisGeNET, NCAM1 gene is associated with Rheumatoid arthritis and Mental disorders such as Bipolar Disorder.

Fas ligand (FASLG) gene is a member of the tumor necrosis factor superfamily. The primary function of the encoded transmembrane protein is the induction of apoptosis triggered by binding FAS. The FAS/FASLG signalling pathway is essential; for immune system regulation, including activation-induced cell death (AICD) of T cells and cytotoxic T lymphocyte induced cell death. It has also been implicated in the progression of several cancers. Defects in this gene may be related to some cases of systemic lupus erythematosus (SLE). Alternatively spliced transcript variants have been described (NCBI). According to DisGeNET, FASLG gene is involved in Rheumatoid arthritis and many other Autoimmune diseases.

CCCTC-binding factor (CTCF) gene is a member of the BORIS + CTCF gene family and encodes a transcriptional regulator protein with 11 highly conserved zinc finger (ZF) domains. It is reported that this nuclear protein can use different combinations of the ZF domains to bind different DNA target sequences and proteins. Depending upon the context of the site, this protein can bind a histone

acetyltransferase (HAT)-containing complex and function as a transcriptional activator or bind a histone deacetylase (HDAC)-containing complex and function as a transcriptional repressor. If the protein is bound to a transcriptional insulator element, it can block communication between enhancers and upstream promoters, thereby regulating imprinted expression. Mutations in this gene have been associated with invasive breast cancers, prostate cancers, and Wilms' tumors. Alternatively spliced transcript variants encoding different isoforms have been found for this gene (NCBI).

The SMAD family member 4 (SMAD4) gene provides instructions for making a protein involved in transmitting chemical signals from the cell surface to the nucleus. The SMAD4 protein is part of a signalling pathway, called the transforming growth factor beta (TGF- β) pathway, that allows the environment outside the cell to affect gene activity and protein production within the cell. The signalling process begins when a TGF- β protein binds to receptor protein on the cell surface, which activates a group of related SMAD proteins. The SMAD proteins bind to the SMAD4 protein and form a protein complex, which then moves to the cell nucleus. In the nucleus, the SMAD protein complex binds to specific areas of DNA where it controls the activity of particular genes and regulates cell growth and division. By controlling these cellular processes, the SMAD4 protein is involved in the development of many body systems. The SMAD4 protein serves both as a transcription factor and as a tumor suppressor (MedlinePlus Genetics). According to DisGeNET, SMAD4 gene is associated with Limitation of joint mobility. This proves that SMAD4 gene is indirectly involved in the onset of Rheumatoid arthritis.

Signal transducer and activator of transcription 1 (STAT1) mediates cellular responses to interferons (INFs), cytokine KITLG/SCF and other cytokines and other growth factors. Following type 1 IFN (IFN-alpha and IFN-beta) binding to cell surface receptors, signalling via protein kinases leads to activation of Jak kinases (TYK2 and JAK1) and to tyrosine phosphorylation of STAT1 and STAT2. The phosphorylated STATs dimerize and associate with ISGF3G/IRF-9 to form a complex termed ISGF3 transcription factor that enters the nucleus (UniProt). Mutations in this gene are associated with Immunodeficiency 31B, 31 A, and 31C (NCBI). According to DisGeNET database, STAT1 gene is involved in

Autoimmune diseases, Skin & Connective Tissue Diseases, Musculoskeletal Diseases and Immune system diseases such as Rheumatoid arthritis, Juvenile pauciarticular chronic arthritis and Polyarticular Juvenile Idiopathic arthritis.

Among the predicted hub genes, EPB42, HBB, SLC4A1, and KLF1 are related to the Hereditary hemolytic anemia; HBB and KLF1 are related to Congenital dyserythropoietic anemia; HBB, AHSP are related to beta-thalassemia; GATA1 is related to X-linked dyserythropoietic anemia and thrombocytopenia (XDAT), Thrombocytopenia with beta-thalassemia, X-linked (XLTT) and Hemolytic anemia due to elevated adenosine deaminase (HAEADA); HYDIN, ZMYDIN and LRRC6 are related to Primary ciliary dyskinesia, Kartagener syndrome and Polynesian Bronchiectasis; STAT1 and CTCF are related to Precursor T-Cell Lymphoblastic Leukemia-Lymphoma; FN1 and NCAM1 are related to Miscarriage, Spontaneous abortion, Early pregnancy loss and Abortion; DNAJC7 is related to Breast carcinoma; FASLG is related to Autoimmune lymphoproliferative syndrome 1A (ALPS1A) and systemic lupus erythematosus (SLE); SMAD4 is related to Juvenile polyposis syndrome and Cholangiocarcinoma (Table 5.4).

These hub genes were further screened to find out the direct association of the predicted genes with Rheumatoid arthritis using DisGeNET database. The results showed that STAT1, FASLG, FN1, NCAM1, and GATA1 genes were directly linked with the Rheumatoid arthritis disease (Table 5.2). Articles such as (Yoshida S et al., 2012) supports the association of STAT1 gene with Rheumatoid arthritis by identifying that the expression levels of signal transducer and activator of transcription 1 is statistically and significantly higher in the synovium of RA than in that of Osteoarthritis; (van der Pouw Kraan TC et al., 2003) proves that there is a prominent role of an activated signal transducer and activator of transcription 1 pathways in rheumatoid tissues; a study conducted by (Zhang R et al., 2018) suggested that increased expression of IL7R and STAT1 in synovial tissues and in the primary immunodeficiency might be associated with the occurrence of RA; (Wang S et al., 2018) has found that E2F2 directly regulates the STAT1 pathways to exacerbate the inflammatory phenotype in rheumatoid arthritis synovial fibroblasts; a research conducted by (Lee HS et al., 2007) have genotyped 67 single nucleotide polymorphisms (SNPs) within the STAT1 and STAT4 regions in 1123 Korean patients with RA and ethnicity-matched controls. Articles supporting

the association of FASLG gene with RA are (Lee YH et al., 2015) which proves the association between FASL rs763110 polymorphisms and rheumatoid arthritis susceptibility through meta-analysis in Asians; and (Yildir S et al., 2013) has found that polymorphisms in the FasL gene related with apoptosis may increase genetic susceptibility to RA in a Turkish population. Articles that support the association of FN1 (Fibronectin 1) with RA are (Yan X et al., 2013), which suggests that PADI4 in the RA synovium may contribute to cartilage destruction via the citrullination of FN; (Silva GL et al., 2009) suggests that the ability of fibronectin fragments to stimulate the mediators of matrix and cartilage destruction in RA paves a way for clearer understanding of the functional role of HLA-DRB1 in RA pathogenesis; (van Beers JJ et al., 2012) proves that 5 citrullinated residues were identified in fibronectin from RA synovial fluid; (Przybysz M et al., 2007) has found the synovial fibronectin fragmentation and domain expressions in relation to rheumatoid arthritis progression. Article that supports the association of NCAM1 gene with RA is (Conigliaro P et al., 2011), which demonstrates that CD56 (bright) Natural Killer cell subset, showing immunoregulatory properties, has been found to accumulate in tissue sites of inflammation, such as the skin lesions in psoriatic patients and the synovial membrane in RA patients. An article which supports the association of GATA1 gene with RA is (Liu H et al., 2018) and the data used in this article suggests that synovial GATA1 is an essential contributor to development and exacerbation of RA, presumably by inducing NOS2 transcription.

We also attempted to identify miRNAs and transcription factors that target the hub genes using the miRNet 2.0 database and obtained a total of 402 miRNAs and 70 TFs as a result (Fig 5.12). From the resulting miRNA-target network, we found the most significant miRNA, hsa-mir-27a-3p, targeting 10 out of 18 hub genes and significant TFs, namely GATA1, NFKB1, and RELA, targeting 6 out of 18 genes. Based on this finding, we can hypothesis that has-mir-27a-3p is an important miRNA, and GATA1, NFKB1, and RELA are important TFs involved in RA pathogenesis and development, and that they might regulate the gene expression in RA. However, further in-vitro and in-vivo research is required to validate our hypothesis. Previous research studies have found that the levels of miR-27a-3p were up-regulated in transforming growth factor- β 1-treated human lung fibroblasts in a Smad2/3-dependent manner and in fibroblasts isolated from

lungs of mice with experimental pulmonary fibrosis (Cui H et al., 2016) and MIR-27a regulates the TGF- β signaling pathway by targeting SMAD2 and SMAD4 in lung cancer (Chae DK et al., 2017).

The results obtained from this study clearly indicates that the identified 10 up-regulated and 10 down-regulated hub genes are directly or indirectly involved in the progression or onset of Rheumatoid arthritis disease.

CHAPTER - 6

SUMMARY AND CONCLUSION

Our study focuses on Rheumatoid arthritis, a serious ailment that affects people all over the world. Based on the results of this study, it can be deduced that there are 5 genes in particular that are more significant, namely STAT1, FN1, FASLG, NCAM1, and GATA1. A few of these genes are also associated with other diseases such as Mental disorders like Bipolar disorder, Anemia, Respiratory disorders and other Autoimmune disorders. In addition, a miRNA, has-mir-27a-3p and three transcription factors, GATA1, NFkB1, and RELA were predicted to play significant role in controlling the gene expression of the hub genes. Therefore, this study indicates that the identified hub genes and miRNAs might provide new concepts for developing diagnosis and serve as a platform for developing therapeutics against Rheumatoid arthritis.

REFERENCES

1. Adam McDermaid, Brandon Monier, Jing Zhao, Bingqiang Liu, Qin Ma, Interpretation of differential gene expression results of RNA-seq data: review and integration, *Briefings in Bioinformatics*, Volume 20, Issue 6, November 2019, Pages 2044–2054, <https://doi.org/10.1093/bib/bby067>
2. Alam, J., Jantan, I., & Bukhari, S.N.A. (2017). Rheumatoid arthritis: Recent advances on its etiology, roles of cytokines and pharmacotherapy. *Biomedicine & Pharmacotherapy*, 92, 615-633. Doi: 10.1016/j.biopha.2017.05.055.
3. Arthritis Foundation. (2019). Arthritis by the Numbers, 3. <https://www.arthritis.org/getmedia/e1256607-fa87-4593-aa8a-8db4f291072a/2019-abtn-final-march-2019.pdf>
4. Bullock J, Rizvi S, A, A, Saleh A, M, Ahmed S, S, Do D, P, Ansari R, A, Ahmed J: Rheumatoid Arthritis: A Brief Overview of the Treatment. *Med Princ Pract* 2018; 27:501-507. doi: 10.1159/000493390
5. Cassotta, M., Pistollato, F. and Battino, M. (2020) “Rheumatoid arthritis research in the 21st century: Limitations of traditional models, new technologies, and opportunities for a human biology-based approach”, *ALTEX - Alternatives to animal experimentation*, 37(2), pp. 223–242. doi:10.14573/altex.1910011.
6. Chae DK, Ban E, Yoo YS, Kim EE, Baik JH, Song EJ. MIR-27a regulates the TGF- β signaling pathway by targeting SMAD2 and SMAD4 in lung cancer. *Mol Carcinog*. 2017 Aug;56(8):1992-1998. doi: 10.1002/mc.22655. Epub 2017 Apr 13. PMID: 28370334
7. Chauhan K, Jandu JS, Brent LH, et al. Rheumatoid Arthritis. [Updated 2022 Nov 21]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK441999/>
8. Cojocaru M, Cojocaru IM, Silosi I, Vrabie CD, Tanasescu R. Extra-articular Manifestations in Rheumatoid Arthritis. *Maedica (Bucur)*. 2010 Dec;5(4):286-91. PMID: 21977172; PMCID: PMC3152850.

9. Conigliaro P, Scrivo R, Valesini G, Perricone R. Emerging role for NK cells in the pathogenesis of inflammatory arthropathies. *Autoimmun Rev.* 2011 Aug;10(10):577-81. doi: 10.1016/j.autrev.2011.04.017. Epub 2011 Apr 21. PMID: 21536152.
10. Crowson, C.S., Matteson, E.L., Myasoedova, E., Michet, C.J., Ernste, F.C., Warrington, K.J., Davis, J.M., III, Hunder, G.G., Thorneau, T.M. and Gabriel, S.E. (2011), The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. *Arthritis & Rheumatism*, 63: 633-639. <https://doi.org/10.1002/art.30155>
11. Cui H, Banerjee S, Xie N, Ge J, Liu RM, Matalon S, Thannickal VJ, Liu G. MicroRNA-27a-3p Is a Negative Regulator of Lung Fibrosis by Targeting Myofibroblast Differentiation. *Am J Respir Cell Mol Biol.* 2016 Jun;54(6):843-52. doi: 10.1165/rcmb.2015-0205OC. PMID: 26600197; PMCID: PMC4942218.
12. Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P. Tsafou, Michael Kuhn, Peer Bork, Lars J. Jensen, Christian von Mering, STRING v10: protein–protein interaction networks, integrated over the tree of life, *Nucleic Acids Research*, Volume 43, Issue D1, 28 January 2015, Pages D447–D452, <https://doi.org/10.1093/nar/gku1003>
13. Derksen, V.F.A.M., Huizinga, T.W.J. & van der Woude, D. The role of autoantibodies in the pathophysiology of rheumatoid arthritis. *Semin Immunopathol* **39**, 437–446 (2017). <https://doi.org/10.1007/s00281-017-0627-z>
14. Fan D, Liu B, Gu X, Zhang Q, Ye Q, Xi X, Xia Y, Wang Q, Wang Z, Wang B, Xu Y and Xiao C (2022) Potential Target Analysis of Triptolide Based on Transcriptome-Wide m6 A Methylome in Rheumatoid Arthritis. *Front. Pharmacol.* 13:843358. doi: 10.3389/fphar.2022.843358
15. Gabriel, S.E., Michaud, K. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther* 11, 229 (2009). <https://doi.org/10.1186/ar2669>

16. Galloway J, Cope AP. The ying and yang of fever in rheumatic disease. *Clin Med (Lond)*. 2015 Jun;15(3):288-91. doi: 10.7861/clinmedicine.15-3-288. PMID: 26031982; PMCID: PMC4953116.
17. Gregersen, P.K., Silver, J. and Winchester, R.J. (1987), The shared epitope hypothesis. an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis & Rheumatism*, 30: 1205-1213. <https://doi.org/10.1002/art.1780301102>
18. Hans Ulrich Scherer, Thomas Häupl, Gerd R. Burmester, The etiology of rheumatoid arthritis, *Journal of Autoimmunity*, Volume 110, 2020, 102400, ISSN 0896-8411, <https://doi.org/10.1016/j.jaut.2019.102400>.
19. Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. doi:10.1038/nprot.2008.211
20. Klareskog, L., Stolt, P., Lundberg, K., Källberg, H., Bengtsson, C., Grunewald, J., Rönnelid, J., Erlandsson Harris, H., Ulfgren, A.-K., Rantapää-Dahlqvist, S., Eklund, A., Padyukov, L. and Alfredsson, L. (2006), A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination. *Arthritis & Rheumatism*, 54: 38-46. <https://doi.org/10.1002/art.21575>
21. Kłodziński, Ł. and Wiśłowska, M. (2018). Comorbidities in rheumatic arthritis. *Reumatologia/Rheumatology*, 56(4), pp.228-233. <https://doi.org/10.5114/reum.2018.77974>
22. Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. *Mol Med*. 2007 Sep-Oct;13(9-10):455-60. doi: 10.2119/2007-00072.Lee. PMID: 17932559; PMCID: PMC2014726.
23. Lee YH, Bae SC, Song GG. Association between the CTLA-4, CD226, FAS polymorphisms and rheumatoid arthritis susceptibility: a meta-analysis. *Hum Immunol*. 2015 Mar;76(2-3):83-9. doi: 10.1016/j.humimm.2015.01.023. Epub 2015 Jan 30. PMID: 25645050.
24. Littlejohn, E. A., & Monrad, S. U. (2018). Early Diagnosis and Treatment of Rheumatoid Arthritis. *Primary Care: Clinics in Office Practice*, 45(2), 237–255. Doi: 10.1016/j.pop.2018.02.010.

25. Lubberts E, van den Berg WB. Cytokines in the Pathogenesis of Rheumatoid Arthritis and Collagen-Induced Arthritis. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK6288/>
26. Ma J, Hu X, Dai B, Wang Q, Wang H. 2021. Bioinformatics analysis of laryngeal squamous cell carcinoma: seeking key candidate genes and pathways. PeerJ 9:e11259 <https://doi.org/10.7717/peerj.11259>
27. Marines du Teil Espina, Giorgio Gabarrini, Hermie J M Harmsen, Johanna Westra, Arie Jan van Winkelhoff, Jan Maarten van Dijk, Talk to your gut: the oral-gut microbiome axis and its immunomodulatory role in the etiology of rheumatoid arthritis, FEMS Microbiology Reviews, Volume 43, Issue 1, January 2019, Pages 1-18, <https://doi.org/10.1093/femsre/fuy035>
28. McWilliams DF, Rahman S, James RJE, Ferguson E, Kiely PDW, Young A, Walsh DA. Disease activity flares and pain flares in an early rheumatoid arthritis inception cohort; characteristics, antecedents and sequelae. BMC Rheumatol. 2019 Nov 18;3:49. doi: 10.1186/s41927-019-0100-9.
29. Mikuls, T.R. (2010), Rheumatoid arthritis incidence: What goes down must go up?. Arthritis & Rheumatism, 62: 1565-1567. <https://doi.org/10.1002/art.27432>
30. Modi, A., Purohit, P., Gadwal, A., Ukey, S., Roy, D., Fernandes, S., & Banerjee, M. (2020, December 2). In-silico analysis of differentially expressed genes and their regulating microRNA involved in lymph node metastasis in invasive breast carcinoma. medRxiv. <https://doi.org/10.1101/2020.11.25.20235259>
31. Palindromic Rheumatism: Longterm Outcomes of 60 Patients Diagnosed in 1967-84. ESSI KOSKINEN, PEKKA HANNONEN, TUULUKKI SOKKA. The Journal of Rheumatology Sep 2009, 36 (9) 1873-1875; DOI: <https://doi.org/10.3899/jrheum.090025>
32. Przybysz M, Borysewicz K, Szechinski J, Katnik-Prastowska I. Synovial fibronectin fragmentation and domain expressions in relation to rheumatoid arthritis progression. Rheumatology (Oxford). 2007 Jul;46(7):1071-5. doi: 10.1093/rheumatology/kem067. Epub 2007 Apr 25. PMID: 17459959.
33. Qiu X, Lin J, Liang B, Chen Y, Liu G and Zheng J (2021) Identification of Hub Genes and MicroRNAs Associated With Idiopathic Pulmonary Arterial

- Hypertension by Integrated Bioinformatics Analyses. *Front. Genet.* 12:667406. doi: 10.3389/fgene.2021.636934
34. Scott, I.C., Machin, A., Mallen, C.D. et al. The extra-articular impacts of rheumatoid arthritis: moving towards holistic care. *BMC Rheumatol* **2**, 32 (2018). <https://doi.org/10.1186/s41927-018-0039-2>
 35. Shannon, P. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504. doi:10.1101/gr.1239303
 36. Silva GL, Junta CM, Sakamoto-Hojo ET, Donadi EA, Louzada-Junior P, Passos GA. Genetic susceptibility loci in rheumatoid arthritis establish transcriptional regulatory networks with other genes. *Ann N Y Acad Sci.* 2009 Sep;1173:521-37. doi: 10.1111/j.1749-6632.2009.04629.x. PMID: 19758195.
 37. Sparks, J. A. (2019). Rheumatoid Arthritis. *Annals of Internal Medicine*, 170(1), ITC1. doi:10.7326/aitc201901010
 38. Tanaka, Y. Rheumatoid arthritis. *Inflamm Regener* 40, 20 (2020). <https://doi.org/10.1186/s41232-020-00133-8>
 39. Udhaya Kumar S, Thirumal Kumar D, Bithia R, Sankar S, Magesh R, Sidenna M, George Priya Doss C and Zayed H (2020) Analysis of Differentially Expressed Genes and Molecular Pathways in Familial Hypercholesterolemia Involved in Atherosclerosis: A Systematic and Bioinformatics Approach. *Front. Genet.* 11:734. doi: 10.3389/fgene.2020.00734
 40. van Beers JJ, Willemze A, Stammen-Vogelzangs J, Drijfhout JW, Toes RE, Pruijn GJ. Anti-citrullinated fibronectin antibodies in rheumatoid arthritis are associated with human leukocyte antigen-DRB1 shared epitope alleles. *Arthritis Res Ther.* 2012 Feb 17;14(1):R35. doi: 10.1186/ar3744. PMID: 22339947; PMCID: PMC3392834.
 41. van der Pouw Kraan TC, van Gaalen FA, Kasperkovitz PV, Verbeet NL, Smeets TJ, Kraan MC, Fero M, Tak PP, Huizinga TW, Pieterman E, Breedveld FC, Alizadeh AA, Verweij CL. Rheumatoid arthritis is a heterogeneous disease: evidence for differences in the activation of the STAT-1 pathway between rheumatoid tissues. *Arthritis Rheum.* 2003 Aug;48(8):2132-45. doi: 10.1002/art.11096. PMID: 12905466.

42. Wang S, Wang L, Wu C, Sun S, Pan JH. E2F2 directly regulates the STAT1 and PI3K/AKT/NF- κ B pathways to exacerbate the inflammatory phenotype in rheumatoid arthritis synovial fibroblasts and mouse embryonic fibroblasts. *Arthritis Res Ther*. 2018 Oct 4;20(1):225. doi: 10.1186/s13075-018-1713-x. PMID: 30286793; PMCID: PMC6235203.
43. Yan X, Yin L, Wang Y, Zhao Y, Chang X. The low binding affinity of ADAMTS4 for citrullinated fibronectin may contribute to the destruction of joint cartilage in rheumatoid arthritis. *Clin Exp Rheumatol*. 2013 Mar-Apr;31(2):201-6. Epub 2012 Nov 7. PMID: 23137648.
44. Yang R, Zhou Y, Du C, Wu Y. Bioinformatics analysis of differentially expressed genes in tumor and paracancerous tissues of patients with lung adenocarcinoma. *J Thorac Dis* 2020;12(12):7355-7364. doi: 10.21037/jtd-20-3453
45. Yannan Fan, Keith Siklenka, Simran K. Arora, Paula Ribeiro, Sarah Kimmins, Jianguo Xia, miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis, *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016, Pages W135–W141, <https://doi.org/10.1093/nar/gkw288>
46. Yıldır S, Sezgin M, Barlas İÖ, Türköz G, Ankaralı HÇ, Şahin G, Erdal ME. Relation of the Fas and FasL gene polymorphisms with susceptibility to and severity of rheumatoid arthritis. *Rheumatol Int*. 2013 Oct;33(10):2637-45. doi: 10.1007/s00296-013-2793-1. Epub 2013 Jun 9. PMID: 23749041.
47. Yoshida S, Arakawa F, Higuchi F, Ishibashi Y, Goto M, Sugita Y, Nomura Y, Niino D, Shimizu K, Aoki R, Hashikawa K, Kimura Y, Yasuda K, Tashiro K, Kuhara S, Nagata K, Ohshima K. Gene expression analysis of rheumatoid arthritis synovial lining regions by cDNA microarray combined with laser microdissection: up-regulation of inflammation-associated STAT1, IRF1, CXCL9, CXCL10, and CCL5. *Scand J Rheumatol*. 2012 May;41(3):170-9. doi: 10.3109/03009742.2011.623137. Epub 2012 Mar 9. PMID: 22401175; PMCID: PMC3400100.
48. Zhang R, Yang X, Wang J, Han L, Yang A, Zhang J, Zhang D, Li B, Li Z, Xiong Y. Identification of potential biomarkers for differential diagnosis between rheumatoid arthritis and osteoarthritis via integrative genome-wide gene expression profiling analysis. *Mol Med Rep*. 2019 Jan;19(1):30-40.

doi: 10.3892/mmr.2018.9677. Epub 2018 Nov 20. PMID: 30483789;
PMCID: PMC6297798.