Genetic Analysis of Rheumatoid Arthritis

To be submitted as Major Project in partial fulfilment of the requirement for the degree of

M. Tech. BIOMEDICAL ENGINEERING

Submitted by

Nisha Gupta

(2K14/BME/08)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. Yasha Hasija

Assistant Professor

Department of Biotechnology

Delhi Technological University, Delhi

CERTIFICATE

This is to certify that the dissertation entitled **"*Genetic Analysis of Rheumatoid Arthritis*"** in the partial fulfilment of the requirements for the reward of the degree of masters in Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by him/her under my guidance. The information and data enclosed in this thesis is original and has not been submitted elsewhere for honouring of any other degree.

Date :

Dr. Yasha Hasija
Assistant Professor
Department of Bio-Technology
Delhi Technological University

Prof. D. kumar
Head Of Department
Department of Bio-technology
Delhi Technological University

ACKNOWLEDGEMENT

I **Nisha Gupta**, student of M.TECH BIOMEDICAL ENGINEERING, 2K14/BME/08 is presenting a project synopsis on "**Genetic Analysis of Rheumatoid Arthritis**" under the supervision of **Dr. Yasha Hasija** (Assistant Professor), department of Biotechnology. She encouraged me to undertake this very interesting topic and even gave me valuable suggestion and information which were mandatory for the completion of the project.

I have taken efforts in this project. However, it would not have been possible without the kind support and help of **Dr. D. Kumar** and organizations. I would like to extend my sincere thanks to all of them.

My thanks and appreciations also go to my colleague in developing and people who have willingly helped me out with their abilities.

Nisha Gupta
2K14/BME/08

DECLARATION

I declare that my major project entitled "**Genetic Analysis of Rheumatoid Arthritis**" submitted to Department of Biotechnology, Delhi Technological University as a result of the work carried out by me at "Genome Informatics Laboratory" Department of Biotechnology, as major project.

Date: 28 /06/2016 Nisha Gupta

Place: Delhi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Genetic Analysis of Rheumatoid Arthritis

Nisha Gupta

Delhi Technological University, Delhi, India

E-mail ID:nisha.biotech2012@gmail.com

ABSTRACT

Current Biomedical research advances in genetics have impelled rapid development towards the systematic identification of genes involved in various disorders. However the detailed understanding of the molecular and physiological mechanisms through which these genes affect disease characteristics remains a major challenge. Genes interact in networks to coordinate cellular processes. Analysis of these networks provides discernments into gene interactions and functions. Here, we will try to identify the Rheumatoid Arthritis disease module, i.e. the local neighborhood of the Interactome whose perturbation is associated with Rheumatoid Arthritis, and validate it for functional and pathophysiological relevance, using computational approaches. We will use different omics data, we will identify the important pathway which is novel modulator in Rheumatoid arthritis. The wiring diagram of the uncovered Rheumatoid Arthritis module suggests a relatively close link between seed genes and other indirectly associated genes. Here, we took advantage of normal variation in human gene expression to deduce gene networks, which we constructed using cytoscape of more than 258 seed genes in Rheumatoid Arthritis. Furthermore, genes that predispose to the same diseases are clustered non randomly in the interaction network, suggesting that networks can provide seed genes that influence disease susceptibility. Therefore, our analysis of gene interaction networks give information on the function of seed genes in normal and disease processes.

CHAPTER 1

INTRODUCTION

There are numerous evidences that genes in both monogenic and complex ailment are not distributed randomly on the molecular Interactome, but they rather work jointly in similar biological pathways. Moreover, gene products related to the same phenotype have a high tendency to interact with each other and also cluster in the same network neighborhood. This advocates the existence of a disease Interactome, a connected sub-network that can be mechanistically related to a particular disease phenotype. The accurate recognition of such disease modules could help unearth the molecular mechanisms of disease causation, identify new seed genes and pathways and aid rational drug target detection.

Currently, the exact identification of disease modules is troubled by the incompleteness of the available cellular network and disease seed genes lists. There is increasing recent advances in Interactome mapping and disease seed gene identification have begun to provide sufficient network coverage and accuracy to enable detection of disease modules for some well-studied complex ailments. The goal of this project is to demonstrate the effectiveness and the value of such an approach, through its significance to the study of Rheumatoid Arthritis, a major complex disease which affect the 1% population of India. Despite the identifications of many susceptibility alleles and genes by GWAS and other technologies, our knowledge of the underlying responsible mechanisms of Rheumatoid Arthritis. Rheumatoid arthritis is the common type of autoimmune arthritis. It is triggered by a faulty immune system and affects mainly the small joints. Treatments have improved very much and help many of those affected. In Rheumatoid arthritis, early treatment can control joint pain, and lessen joint damage. Expertise is vital to make an early diagnosis of Rheumatoid arthritis and to rule out diseases that mimic Rheumatoid arthritis, thus avoiding unneeded tests and treatments.

Rheumatoid arthritis is a chronic disorder that mainly affects joints but In some cases, the condition also can harm a wide variety of body systems, including the skin, eyes, lungs, heart and blood vessels. An autoimmune disorder, rheumatoid arthritis occurs when the own immune system mistakenly attacks the body's tissues not the wear-and-tear damage. Rheumatoid arthritis affects the joints, causing a painful swelling that can eventually result in joint deformity.

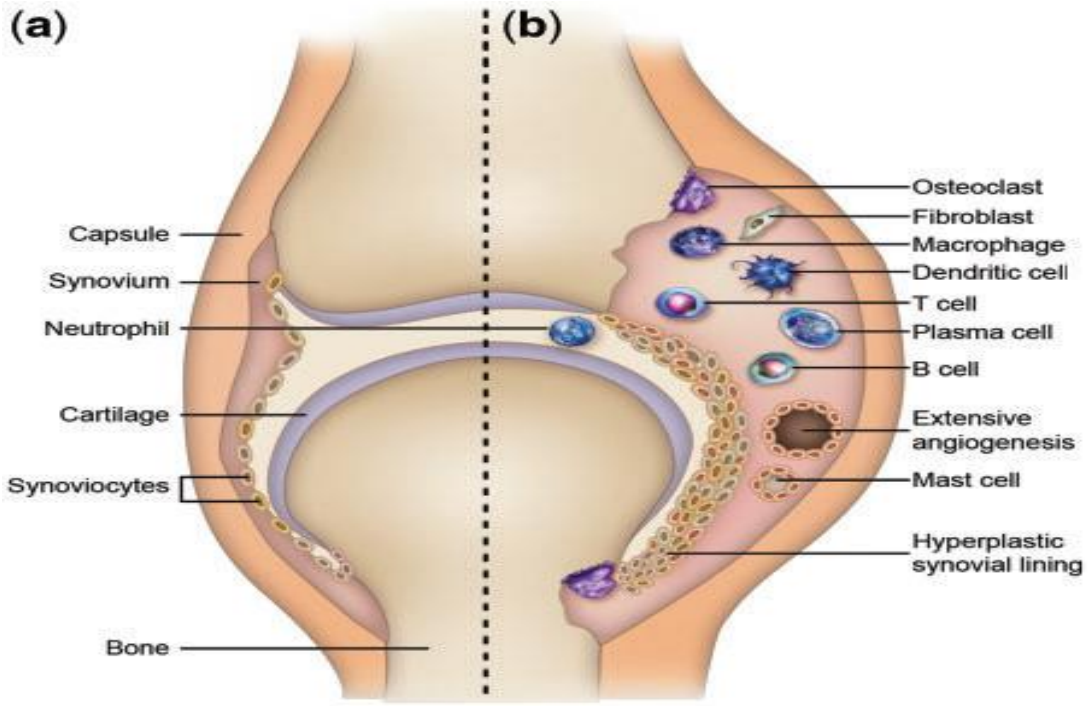


Figure 1.1 Representation of normal joint and joint affected with Rheumatoid Arthritis.

The inflammation linked with rheumatoid arthritis is what can damage other organs as well. While new types of treatment have improved options, severe rheumatoid arthritis can still cause physical disabilities.

1.1Causes

The condition of Rheumatoid arthritis arises when immune system attacks the synovium — membranes that surround joints. The affect of inflammation thickens the synovium, which can destroy the cartilage and bone of the joint. The tendons and ligaments which hold the joint together deteriorate and stretch. Gradually, the joint loses alignment. Doctors can not recognize what starts this process, although a genetic component appears likely. While genes don't actually cause rheumatoid arthritis, they can make the person more susceptible to environmental factors — such as infection with certain pathogens— that may trigger the disease.

1.2 Risk factors

Factors that may increase the risk of rheumatoid arthritis include, Sex, Age, Family history, Smoking and Environmental exposures.

1.3 Complications

Rheumatoid arthritis enhances your risk of developing Osteoporosis, Rheumatoid nodules, Dry eyes, mouth Infections, Abnormal composition of body, Carpal tunnel syndrome, Heart problems, Lung disease, Lymphoma.

1.4 Diagnosis of Rheumatoid Arthritis

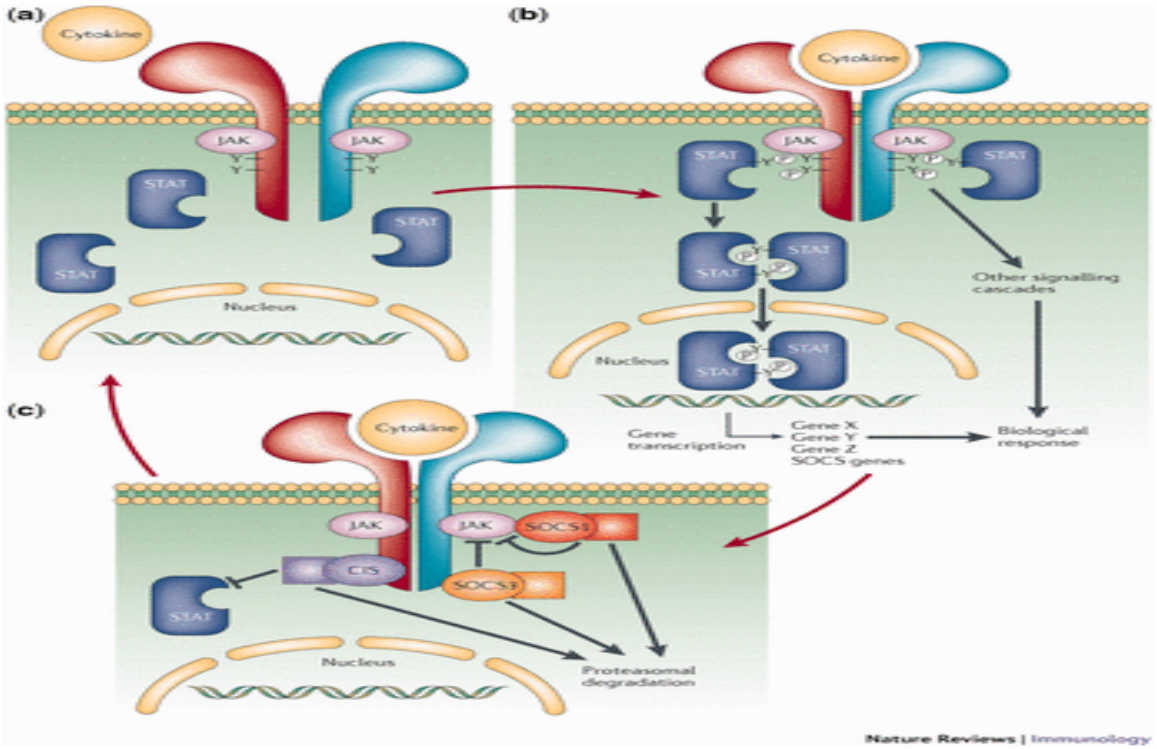
Rheumatoid Arthritis can be tough to detect because earlier symptoms of this is similar to other joints ailments such as achy joints or a little stiffness. Diagnosis of RA depends on the symptoms and results of a physical exam. Some blood tests also can help confirm Rheumatoid Arthritis. There is no any single test that confirms Rheumatoid Arthritis diagnosis for most patients with this disease.

1.5 Genetics of Rheumatoid Arthritis

Rheumatoid Arthritis can be characterized by the infiltration of various inflammatory cells into joint fluid, and later tissue destruction. Imbalance concerning pro- and anti-inflammatory cytokine activities supports the initiation of autoimmunity, chronic inflammation and thereby damage of joint. The macrophages play a pivotal role in Rheumatoid Arthritis because they are several in the inflamed synovial membrane. They activates MHCII molecule, and secretes pro-inflammatory and growth factors like IL1, IL2, IL6, IL10, IL13, IL15, IL17, IL18, TNFα ,GM-CSF Chemokines and Chemo-attractants Metalloproteinases. Expression of IL1β regulated by TNF, which is important for the induction of prostanoid and Metalloproteinases production by synovial fibroblasts and chondrocytes. TNF enhances the expression of adhesion molecules on endothelial cells, which recruit number of cells to the joint. IL1 and TNF induce synovial fibroblasts to express IL6, chemokines, GM-CSF and MMPs, which take part in cartilage and bone destruction. TNF contributes to osteoclast activation and differentiation. In addition, IL1 mediates degradation of cartilage directly by inducing the expression of Metalloproteinases by chondrocytes. In addition, these chemokines, produced by Rheumatoid Arthritis synovial stromal cells also stimulate monocyte migration. Other cytokines such as IFNγ induced chemokines also contribute to the documented morphologic and clinical features of Rheumatoid Arthritis. Autoreactive B cells can be driven by the T cells to produce IgG autoantibodies that may be directly involved in joint damage. As the B cell appears to play an essential role in the Rheumatoid Arthritis process, it is appropriate to consider how B cell-mediated effects which might be reduced in patients with this disease.

Rheumatoid Arthritis also caused by abnormal presentation of self antigen by APCs and activation of autoreactive T-Cells. T lymphocytes play a main role in the disease process. APCs require signals from activated T-Cells for differentiation and maturation; this subsequently facilitates APCs to activate newly arrived T-Cells in a specific or unspecific manner in inflammation. Several co-stimulatory molecules are involved at the time of APC T-Cell interactions, includingCD28/CD80-CD86 andCD40-CD40L. Some of these molecules are crucial in initiation of the immune response CD28/CD80/CD86,whileCD40-CD40L is required for the enlargement of the inflammatory response. Huge numbers of activated leukocytes infiltrate the synovial membrane, causing inflammation, In most of the cases which lead to progressive destruction of cartilage and bone. Since Rheumatoid Arthritis is a systemic autoimmune disease, other parts/organs of the body may become affected at a delayed stage. Evidence of environmental factors, such as infectious agent and smoking, may take part a role. Although the contributing mechanisms of the pathogenesis of RA are unknown, a genetic predisposition has been identified in certain groups. This genetic predisposition, and also the activation and affinity maturation of autoreactive T-Cells and B-Cells that are in the joint, point out a role for adaptive immunity in the pathogenesis of RA.

1.6 Role Of JAK STAT Pathway In Rheumatoid Arthritis



Kelly *et. al.*

Figure1.2 Activation of JAK STAT signaling pathway by cytokines

The family of non receptor protein tyrosine kinases are JAKs that affect intracellular signalling in association with transcription factors known as STATs, otherwise known as the JAK–STAT pathway. JAKs are bound to their receptors and then activated when the subsequent cytokine or growth factor binds to that particular receptor. Activated JAKs phosphorylate the tyrosine residues on the receptor and then causes a change in conformation, allowing the binding of a STAT protein in the SH2. JAKs then phosphorylates the residues of tyrosine on the STAT protein, allowing the dimerization of the STATs, which afterwards migrate into the cytoplasm and translocate into the nucleus, permitting for transcription of target genes. JAKs also autophosphorylate, but the involvement of this is not well understood till date.

1.7 Treatment of Rheumatoid Arthritis

There is no cure for Rheumatoid Arthritis. The goal of treatment is to minimize the symptoms and poor function. Good control of Rheumatoid Arthritis requires earlier diagnosis. Thus, patients with a diagnosis of Rheumatoid Arthritis should begin their treatment with disease-modifying anti-rheumatic drugs, DMARDs. These drugs not only reduce symptoms but also slow progression of the disease. DMARDs have greatly improved the symptoms, function and life quality for nearly all patients with Rheumatoid Arthritis.

1.8 Living with rheumatoid arthritis

Research shows that people with RA, mainly those whose disease is not well controlled, have a higher risk for heart disease and stroke. Talk with your doctor about these risks and ways to lower them.

It is important to be physically active most of the time, but to sometimes scale back activities when the disease flares. In general, rest is helpful when a joint is inflamed, or when you feel tired. At these times, do gentle range-of-motion exercises, such as stretching. This will keep the joint flexible.

Traditional single gene or single pathway based approaches have shown limited utility. In the past few years, several attempts have been made to integrate the topological properties of protein interaction networks with different types of ‘omics’ data to discover novel genes and pathways. These approaches rely on the local impact hypothesis, assuming that if a few disease components are identified, other disease-related components are likely found in their network vicinity. Therefore, each disease can be linked to a well defined local neighborhood

of the Interactome, called the disease module. Yet, the existence of such a single neighborhood remains a hypothesis that needs to be tested. Our goal, here, is to determine whether a whole network based approach can enhance our understanding of the local network neighborhood of a disease using Rheumatoid Arthritis as an example. We start by compiling a list of physical and experimentally documented interactions in human cells from the literature, as well as a set of known and well-established disease genes. These seed genes allow us to pinpoint the position of the putative disease module within the interactome (Stage I). Next, we apply a network theoretic procedure to identify the localization of potential Rheumatoid Arthritis genes that may belong to this disease module (Stage II). In Stage III the obtained disease module is validated for functional and pathophysiological relevance using several Rheumatoid Arthritis specific biological datasets. We further explore the overlap and the difference between the Rheumatoid Arthritis module and other immune related disease modules. Finally, in Stage IV we explore the pathways within the module that contain promising therapeutic targets.

1.9 Role of SNPs

Genetic analysis is the overall process of studying and researching in fields of science that involve genetics and molecular biology. There are a number of applications that are developed from the various researches, and they are also considering as part of the various process. The basic system of analysis revolves around genetics studies include identification of genes and inherited disorders.

1.10 Role of Protein Protein Interaction Network

Modelling, analyzing and interpreting protein-protein interaction (PPI) networks are an important problem in systems biology. Many random network models were designed to capture particular network properties or mimic the way real PPI networks might have evolved. Thus, our new approach allows us to utilize high quality parts of currently available PPI data to create accurate models for PPI networks of different species.

The majority of cellular functions are not controlled by single proteins, but in association of proteins acting together. There are huge amount of data available which requires new mathematical and computational approaches to be developed for model and analyze the complex networks they form. A suitable model of PPI networks will allow better estimating all types of statistics as well as generating synthetic networks of groups for which protein-

protein interactions have not been determined experimentally. This could help to understand cellular processes and future designing biological experiments. Therefore, analyzing and modelling PPI networks has become an exciting research area.

Many models of network have previously been introduced to capture particular sets of PPI network properties, or to mimic the way in which these models might have evolved. However, it is difficult to say to what extent network properties really describe network structure. Therefore, different models of network should use all of the available real network models information, i.e., the entire network connectivity information. Using our network and our previous interpretations that geometric random graphs provide a model for PPI networks.

CHAPTER 2

LITERATURE REVIEW

Rheumatoid Arthritis is a type of autoimmune disorder of joints and this ailment distributed across all over the world. Plenty of research work has been done for finding out the factors which is responsible for the diseased state. Various research work that has been already done in finding out factors causing this disease . The some significant research work survey which is related to my current research is as follows-

The role of bioinformatics is essential in both interpreting genomic, transcriptomic and proteomic data generated by high throughput experimental technologies. This is base to provide both conceptual and practical methods for detecting systemic functional behaviours of the cell and the organism. (Kanehisa *et. al.*).

Interactions are the essence of all biomolecules because they cannot fulfil their roles without interacting with other molecules. Hence, mapping the interactions of biomolecules can be useful for understanding their roles and functions. In recent years, the mapping of protein-protein interactions in different species has been reported, but few reports have focused on the large-scale mapping of protein-protein interactions in human. In this, developments in protein interaction mapping and discuss issues and strategies for the mapping of the human protein Interactome (Figeys *et.al.*).

2.1 Database and Software

Cytoscape is a freely available software package for visualizing, modelling and analyzing molecular and genetic interaction networks. Cytoscape use to analyze the results of mRNA expression profiling, and other functional genomics and proteomics data, in the context of an interaction network obtained for genes of interest.(Cline *et. al.*). Analysis of network interactions between genes and proteins has become a main theme in systems biology. Multipurpose software tools for visualizing and analyzing these networks are therefore very much in demand. It is an open software environment Cytoscape has been developed with the goal of facilitates the designing and development of such software tools by the scientific community (Vlasblom *et. al.*). A different mode of genetic interaction deciphers different functional relationships between genes. The mining of biological information from multi-mode genetic-interaction networks demands suitable statistical and computational methods.

Development of such methods and implemented them in open-source software. Motifs extracted from multipurpose network of genetic interaction form functional subnetworks, highlight genes dominating these subnetworks, and expose genetic reflections of the underlying biochemical system.(Taylor *et. al.*). Comparision of metabolic important tissue and then ontology study.(Nakai *et. al.*).The BiNGO is an open-source Java tool to determine which Gene Ontology (GO) terms are considerably overrepresented in a set of genes. BiNGO plugin can be used either on a list of genes, text, subgraphs of networks visualized in Cytoscape. BiNGO plugin maps the predominant functional themes of the gene set on GO hierarchy, and takes advantage of Cytoscape's visualization environment to create an insightful and customizable visual representation of the results.(Maere *et. al.*).

The huge amount of protein sequence data are available, for the modelling of protein sequence and function, The PANTHER database was constructed to model evolutionary sequence–function relationships. There are a number of applications for these data as, protein classification service, expression data analysis service, and coding single nucleotide polymorphism scoring service.(Thomus *et. al.*).

2.2 Network development and essence of network

The integration of biomedical data and further use for drug discovery is remaining a challenge. Biological processes such as metabolic pathways, protein-protein interactions are often represented as network in systems biology. The understanding of such networks and their analysis are today important challenges in life sciences. While a great variety of visualization tools that try to deals most of these challenges already exists, only few of them become successful to bridge the gap between visualization and network analysis (Pavlopoulos *et. al.*).

Gene expression profiling and the analysis of PPI networks may support the identification of disease bio-markers and potential drug targets. Thus, a step forward in the development of systems approaches to medicine is the integrative analysis of these data sources in specific pathological conditions. This investigation supports the hypothesis that the integrative analysis of differential gene expression and PPI network analysis may facilitate a better understanding of functional roles and the identification of potential drug targets in human heart failure (Camargo *et. al.*).

The main understanding of the causes of human disease can come from classifying characteristics that are specific to disease genes. However, a full understanding of the role of important genes to human disease is lacking, due to the premise that these genes have a propensity to cause developmental abnormalities rather than adult disease (Dickerson *et.al.*).

SNPs data extracted from GWAS are popular strategies to reveal the genetic basis of human complex diseases. Despite many successful analysis of GWAS, it is well recognized that new analytical approaches have to be integrated to attain their full potential. The characterization of the interrelating behaviours of complex biological systems is a basic objective of protein–protein network analysis and computational biology. Undirected protein– protein networks of interacting proteins intimately correlated with significant biological pathways.(Natale *et.al.*). The normal variation in human gene expression to infer gene networks, which we created using correlations in expression levels of gene in cells from independent samples. The resulting networks allowed us to classify biological processes and gene functions. Furthermore, genes that predispose to the alike diseases are making cluster non-randomly in the co-expression network, suggesting that networks can provide genes that influence ailment susceptibility. Therefore, our analysis of gene co-expression networks offers information on the function of human genes in normal and disease processes (Nayak *et. al.*). The specific SNPs that account for this enrichment can be used as a basis for focused genotype-phenotype studies (Freudenberg *et.al.*).

In current research, various analysis were performed and try to develop the disease module of rheumatoid arthritis.

CHAPTER 3

RESEARCH PROPOSAL

3.1 Introduction

In current biomedical research there is increasing facts that seed genes in diseases are not interact randomly on the molecular interaction network i.e. Interactome, but they work together in similar biological modules. In addition, gene products linked to the same phenotypic characters have a strong tendency to interact with each other and form cluster in the same network neighborhood. This implies the existence of a disease module, a connected sub-network that can be mechanistically related to a particular disease phenotypic character. The accurate detection of such disease modules could help to decipher the molecular mechanisms of disease causation, identify new disease seed genes and pathways and assist rational drug target detection.

3.2 Theoretical framework and hypotheses

Hypothesis of this research is as, there are various genes which is involve to generate the particular ailment condition. there are number of genes which is directly involve in diseased state known as seed genes but all those genes which are not directly involved in diseased condition but play important role in that condition. In this particular research we are try to find out the disease genetics and alteration in that by using number of tools and find the number of parameter as SNP analysis, pathway analysis, ontology analysis, functional analysis, PPI analysis.

3.3 Methods

3.3.1 Data collection

Various attributes define for the data collection

a) Disease
b) Polymorphic genes
c) PMID
d) Rs id
e) Population
f) P value

g) Odds ratio
h) RNA
i) DNA
j) Chromosome number
k) Chromosomal location

3.3.2 Selection of tools for analysis

3.3.3 Network construction

3.3.4 Network Analysis of data using various parameters

1. **SNP analysis**
2. **Ontology analysis**
3. **Functional analysis**
4. **Pathway analysis**
5. **PPI analysis**

3.4 Research goals and specific aims

The major goal of this research was to develop an Interactome of rheumatoid arthritis using bioinformatics approaches for the analysis of genes, pathways and ontology.

3.5 Specific aims

- Analysis of the shared SNPs/genes associated with rheumatoid arthritis
- Validation of associated SNPs/genes (seed genes) characterize different pathways to rheumatoid arthritis
- Hub gene identification
- To develop an Interactome for analysis of rheumatoid arthritis.
- Gene Ontology Analysis

3.6 Result and discussion

A. Network analysis using Cytoscape

1-Identification of associated genes association

2-Validation of associated genes

3-Hub gene identification

B. Gene Ontology Analysis

C. Gene functional analysis

D. Pathway analysis

E. PPI analysis

CHAPTER 4

RESEARCH DESIGN AND METHODOLOGY

4.1 DATA COLLECTION

To analyze the seed genes of Rheumatoid Arthritis and SNP information was extracted from the central PUBMED server. Thousands of genes reported which show direct or indirect association with the Rheumatoid Arthritis among them various genes shows polymorphisms. In current research we collected 258 polymorphic genes which is found to be positively reported and accessible. Information regarding genes polymorphism were collected and used to develop a data sheet which is based on the following attributes.

Attributes: Disease Name, Gene Name, PUBMED ID, p value, Odd Ratio, Polymorphisms, Population, Chromosome Number, Location, RS-ID, Variant DNA, Variant RNA.

1	Disease name	GeneName	PUBMED-ID	p value	Odd ratio	Polymorphism	Population	Chromosome	Location	RS-ID	Varient D	Varient RNA
2	Rheumatoid A	BGLAP	22842327	0.002	1	C/T	European	1	156240066	rs154329	NC_00000	NM_001199653.1:c.*479C>T,NM
3	Rheumatoid A	C1QA	23607884	0·0006	0·72	A/G	North America	1	22638465	rs292001	NC_00000	NM_015991.2:c.164-368G>A
4	Rheumatoid A	C1QB	23607884	0·0006	0·72	A/G	North America	1	22638465	rs292001	NC_00000	NM_015991.2:c.164-368G>A
5	Rheumatoid A	C1QC	23607884	0·0006	0·72	A/G	North America	1	22638465	rs292001	NC_00000	NM_015991.2:c.164-368G>A
6	Rheumatoid A	CD2	23261300	<0.05		C/G/T	European and	1	116768525	rs624988	NC_00000	NM_001767.3:c.798C>A
7	Rheumatoid Art	CD244	21586211	.	.	A/G	European	1	160839213	rs668265	NC_00000	NM_001166663.1:c.671-164C>
8	Rheumatoid A	CD247	23861880	0.33	0.97	C/T	European Cauc	1	167442147	rs864537	NC_00000	NM_000734.3:c.59-1380T>C,N
9	Rheumatoid A	CD58	19898481	0.001	.	C/G	European des	1	116720516	rs115862	NC_00000	1.10:g.117263138C>G,NC_00000
10	Rheumatoid A	CD84	23555300	0.004	.	A/G	European and	1	160546518	rs642752	NC_00000	NM_001184879.1:c.*1738T>C,NI
11	Rheumatoid A	CFH	19567623	< 0.05	.	C/T	Spanish ances	1	196690107	rs106117	NC_00000	NM_000186.3:c.1204C>T,NM_0
12	Rheumatoid A	CHI3L1	20300754	.	.	C/G	Hungarian	1	203186754	rs495092	NC_00000	NM_001276.2:c.-131C>G
13	Rheumatoid A	CHRM3	21450750	0.0033	1.93	C/T	Swedish and N	1	239835325	rs754852	NC_00000	NM_000740.2:c.-20+7947C>T
14	Rheumatoid A	CNR2	25974389	0.001	.	CC/TT	Italian children	1	23875429	rs357613	NC_00000	NM_001841.2:c.188_189delAAin
15	Rheumatoid A	CRP	25359432	< 0.05	.	C/T	Region Midtjy	1	159699194	rs112652	NC_00000	1.10:g.159668984C>T,NC_0000
16	Rheumatoid A	FASLG	23285481	<0.03	.	C/T	.	1	172658358	rs763110	NC_00000	NM_000639.2:c.-844C,NM_00
17	Rheumatoid A	FCGR2A	26314337	.	1.58	C/T	European and	1	161509955	rs180127	NC_00000	NM_001136219.1:c.500A>G,NM
18	Rheumatoid A	FCGR3A	26314337	0.006	1.21	C/G/T	European and	1	161544752	rs396991	NC_00000	NM_000569.6:c.634T>G,NM_0
19	Rheumatoid A	FCGRL3	25566937	.	.	A/G	Netherlands, S	1	157701026	rs752868	NC_00000	NM_052939.3:c.-461T>C,
20	Rheumatoid A	FOXJ3	23933633	0.05	.	C/T	.	1	42335634	rs585320	NC_00000	NM_001198850.1:c.-26G>A,NM
21	Rheumatoid A	IL10	25623518	.	.	A/C	East Chinese	1	206773062	rs180087	NC_00000	NM_000572.2:c.-627A>C,
22	Rheumatoid A	IL23R	25858864	0.562	.	A/G	Egyptian patien	1	67240275	rs112090	NC_00000	NM_144701.2:c.1142G>A,
23	Rheumatoid A	IL6R	24668548	0.04	.	C/T	chinese	1	154454494	rs668443	NC_00000	NM_000565.3:c.86-5833C>T,N
24	Rheumatoid A	MIA3	22577832	0.018	.	A/C	Spanish	1	222650187	rs174656	NC_00000	NM_001300867.1:c.266-105A>C
25	Rheumatoid A	MMEL1	22365150	< 0.0038	.	A/G	European, Afric	1	2622185	rs389074	NC_00000	NM_033467.3:c.154+7146A>G
26	Rheumatoid A	MTHFR	26314492	0.057	3.15	C/T	Italy and Hunga	1	11796321	rs180113	NC_00000	NM_005957.4:c.665C>T,
27	Rheumatoid A	NLRP3	26440629	0.045	0.6	C/G	European Leag	1	247448734	rs107545	NC_00000	NM_001079821.2:c.*230G>C,NI
28	Rheumatoid A	PADI4	26546893	0.006	1.63	A/G	Iranian populati	1	17336167	rs174803	NC_00000	NM_012387.2:c.349T>C

Table 4.1 Collection of data using the suitable attribute

4.2 BIOINFORMATICS DATA SOURCES UTILIZED IN THE CASE STUDY

S.No.	Tools	Analysis	Origin	URL
1	PANTHER	Ontology study	Version 10.0	http://www.pantherdb.org/
2	CYTOSCAPE	Interaction network	Cytoscape 3.4.0	http://www.cytoscape.org/
3	DAVID	Functional	DAVID Bioinformatics resources 6.7	https://david.ncifcrf.gov/home.jsp
4	STRING	PPI	Version 10.0	http://string-db.org/

Table 4.2 Software and Interface used for the Result analysis

4.3 SNP ANALYSIS

For the all selected gene set SNP were analyzed on the basis of single nucleotide polymorphism associated to Rheumatoid Arthritis using various attributes as chromosome number, chromosomal location.

4.4 DEVELOPING INTERACTOME

4.4.1Cytoscape

Data sources, import and update

The File menu contains most basic file functionality:

1-File Import for importing data such as networks and attributes,

2-File Export for exporting data and images,

3-File Print allows printing,

4-File Quit closes all windows of Cytoscape and exits the program,

5-File New creates a new network, either blank for editing, or from an existing network,

6-File Open for opening a Cytoscape session file,

7-File Save for saving a session file,

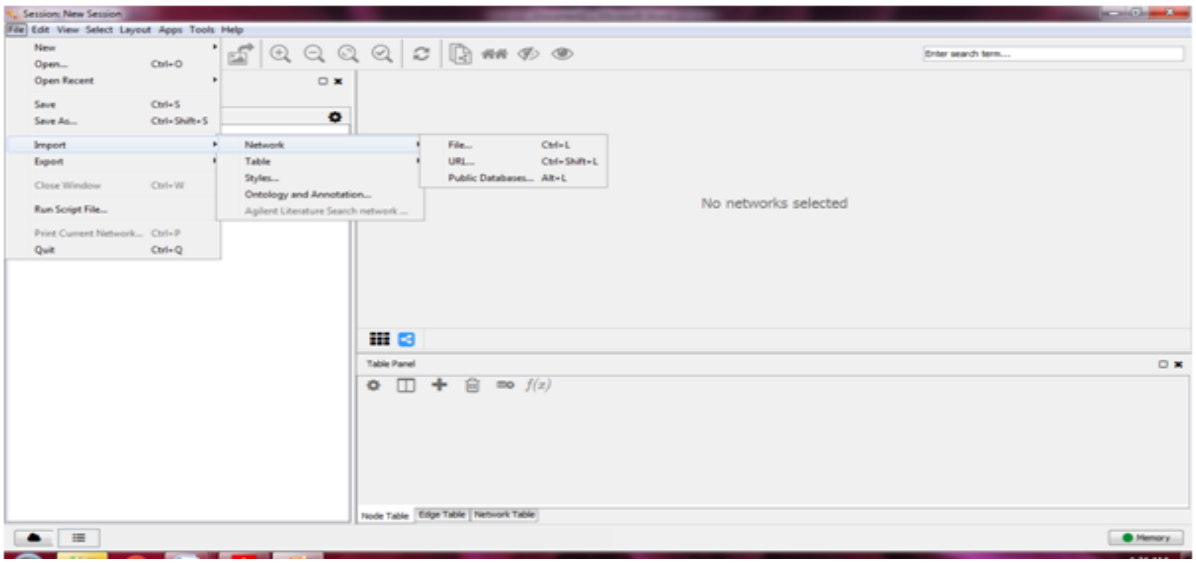


Figure 4.1 Import for importing data such as networks and attributes in XLS format.

Interaction data files

Attribute Names- genes, p value etc.

Transfer first line as attribute names: By choosing this option, first entry in the file will be used as edge attribute names.

Network Import Options: If Interaction Type is set to **Default Interaction**, the value here will be used as interaction for all edges.

We can add arbitrary node, edge and network information to Cytoscape as node/edge/network

Advanced options

These attributes can then be visualized in a user-defined way by setting up a mapping from data to visual attributes (colors etc.).

Zoom Cytoscape provides two mechanisms for zooming, either using mouse gestures or buttons on the toolbar. Use the zooming keys located on the toolbar to zoom in / out of the interaction network shown in the current display.

4.5 GENE ONTOLOGY ANALYSIS

4.5.1 GOrilla analysis

For ontology analysis we are providing genes which are obtained from the literature survey. Literature survey involves capturing published information about the gene and its product as GO annotations. In the given set of genes that are up-regulated under certain conditions, an

enrichment analysis will find which GO terms are represented (over or under) using annotations for that gene set.

Below is a schematic diagram giving an introduction to the steps involved in literature-based GO analysis as:

Figure 4.2 First, select the name of the species. Second, choose the running mode. Third, paste list of the genes to be analyzed, one per row. The ontology which we want to calculate the enrichment (biological process, molecular function, or cellular component). To minimize search time, the tool searches only one ontology at a time and also can calculate collectively at a time.

GOrilla is publicly available as a web interface application. This application can be operated using two modes

1. Discovery of enriched GO terms for the top ranked list of genes using the mHG statistics.

2. Discovery of enriched GO terms in a target set versus a background set and using a hyper geometric model

3. GOrilla automatically removes duplicates keeping the highest level occurrence. This includes dealing with duplicates that hide behind different nomenclatures.

4. GOrilla at present supports the following organisms: human, mouse, rat, yeast, D. *melanogaster*, C. *elegans* and A. *thaliana*. The output consists of a colour-coded trimmed DAG of all significantly enriched GO terms.

5. The results page displays a flow chart and a table that lists used to describe the set of genes that entered on the, the GO term, Description, p value, FDR q value, Enrichment (N, B, n, b) the sample frequency, Expected p-value. In addition, the results page displays all the criteria used in the analysis.(Eden *et. al.*)

The resulting enriched GO terms is visualized using a DAG graphical representation with colour coding reflecting their degree of enrichments. Nodes in the graph are clickable and give additional information on the GO terms and genes attributing to the enrichment. N is the total number of genes; B is the total number of genes associated with a specific GO term; n is the flexible cutoff, i.e. the automatically determined number of genes in the 'target set' and b is the number of genes in the 'target set' that are associated with a specific GO term. Enrichment is defined as (b/ n)/(B/N).

PPI ANALYSIS: String

The STRING database aims to provide such a global perspective for as many organisms as feasible. Known and predicted associations are scored and integrated, resulting in comprehensive protein networks. Below is a schematic diagram giving an introduction to the steps involved in STRING analysis as:

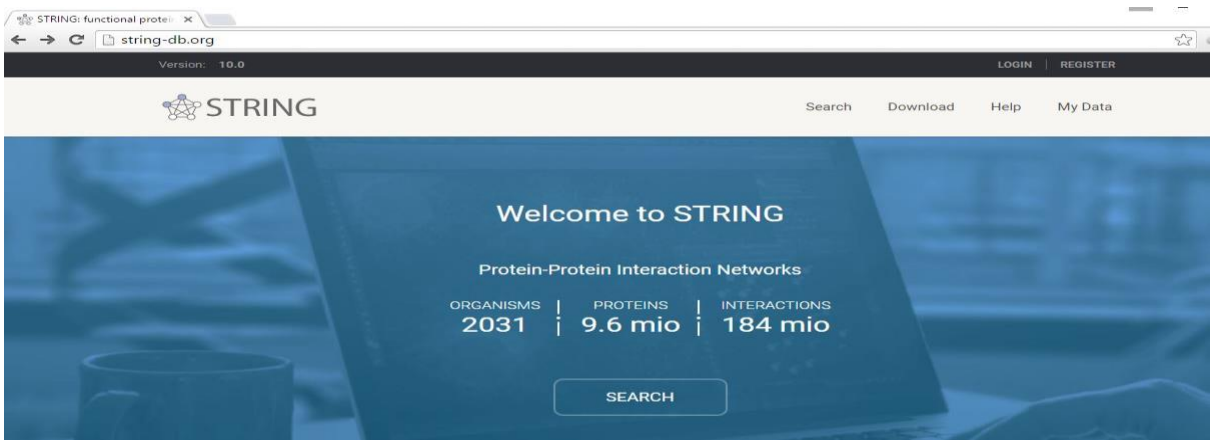


Figure 4.3 First, paste the name of genes one per row. Second, choose the running mode. Third, select the organisms.

Functional Analysis by DAVID

The large gene lists which derived from literature mining is important foundations that directly influences the success of functional analysis in DAVID. Due to the complexity of the data mining situations involved in biological studies, there is no good systematic way, at the present time, to quantitatively estimate the quality of the gene list ahead of time. However, based on real-life data analysis, the gene list contains all of following characteristics:

1) Contain many important genes (marker genes, seed genes) as expected for the rheumatoid arthritis
2) Reasonable number of genes ranging from 100-1000.
3) Most of the genes significantly pass the statistical threshold (e.g. P-values <=0.05)

The foundation of enrichment analysis is that if a biological process is abnormal in a given study, the co-functioning genes should have a greater potential (enriched) to be selected as a relevant group by the high-throughput technologies. To decide the degree of enrichment, a certain background must be set up in order to perform the comparison.

Upload gene list, Submit to David system using default options and View result in different different mode.

Figure 4.4 Submit a gene list to DAVID and access various analytic tools/modules.

a. Following the example input format and steps on the left side uploading panel, a list of genes may be uploaded into DAVID. **b.** After successful uploading of gene list, a set of

analytic modules are available for the analysis of the current gene list highlighted in the gene list manager on the left side.

This is a useful analytic module particularly when we want to closely look at the annotation of highly interesting genes.

(A) Click on the gene name to link for more detailed information.
(B) Click on "RG" (related genes) beside the gene name to search for other functionally related genes.
(C) Use the browsers "Find" function to search for particular items.

(D) Click on the red "G" to list all associated genes of all terms within the cluster.

(E) Click on the "green icon" to display the 2-D (gene-to-term) view for all genes and terms within the cluster.

The annotation terms are clustered using the default clustering stringency. this may re-run the classification function leading to optimal results for the particular case by resetting the stringency (high, medium or low) in the options on top of the result page.

- Visualize genes on BioCarta and KEGG pathway map.
- Display related many genes to many terms on 2D view.
- List Interacting proteins.
- Explore gene name in batch.
- Link gene-disease associations.
- Highlight protein functional domains and motifs.
- Redirect to related literatures.
- convert gene identifiers from one type to another
- Diagnose and fix problems with gene IDs
- Explore gene names in batch
- Highlight functional domains and motifs
- Redirect to related literature
- Cluster redundant and heterozygous annotation terms
- Read the entire annotation contents associated with a gene
- Identify biological themes, particularly GO terms.
- Discover enriched functionally-related gene groups.

CHAPTER 5

RESULT AND DISCUSSION

The Development of Interactome for determining novel pathways and genes of Rheumatoid Arthritis we had taken polymorphic genes and their respective parameters. we obtain the result for the analysis of genes related to Rheumatoid Arthritis some of the essential features which are used to get the desired results are **Name of the gene , RSID , PMID , p-value and odds ratio**. The analysis of current research is as follows:

5.1 SNP ANALYSIS

For the selected gene set all the SNP were analyzed and it was found that maximum of the single nucleotide polymorphism associated to Rheumatoid Arthritis were reported in intronic region. It was found that all the SNPs were form the locus of almost all the chromosomes (except Y).

5.2 DEVELOPING INTERACTOME: HUB GENE IDENTIFICATION

This is very surprising property of biological networks is the appearance of hub proteins, suggesting that hubs must play a unique role. Indeed, evidence from model organisms indicates that hub proteins tend to be encoded by seed genes, and that genes encoding hubs are must be older and evolve more slowly than genes encoding non-hub proteins. The deletion of genes encoding hubs also leads to a enormous phenotypic outcomes than the deletion of genes encoding less connected proteins. It is assume that in humans, hubs should be associated with disease genes. it was found that seed genes show a strong tendency to be linked with hubs and expressed in multiple tissues, i.e., they tend to be located at the functional centre of the Interactome. That is, from a network perspective, these genes segregate at the functional periphery of the Interactome. Most nodes have the same number of links, and maximum connected nodes (hubs) are quite rare in a random network. The fraction of bond with a given degree, called the degree distribution, follows the well-known Poisson distribution. The most noticeable consequence of this property is the presence of a few maximum connected hubs that hold the whole network together. The biological role and dynamical behaviour of hubs are function inside modules and coordinate specific cellular processes, different processes and organize the Interactome.

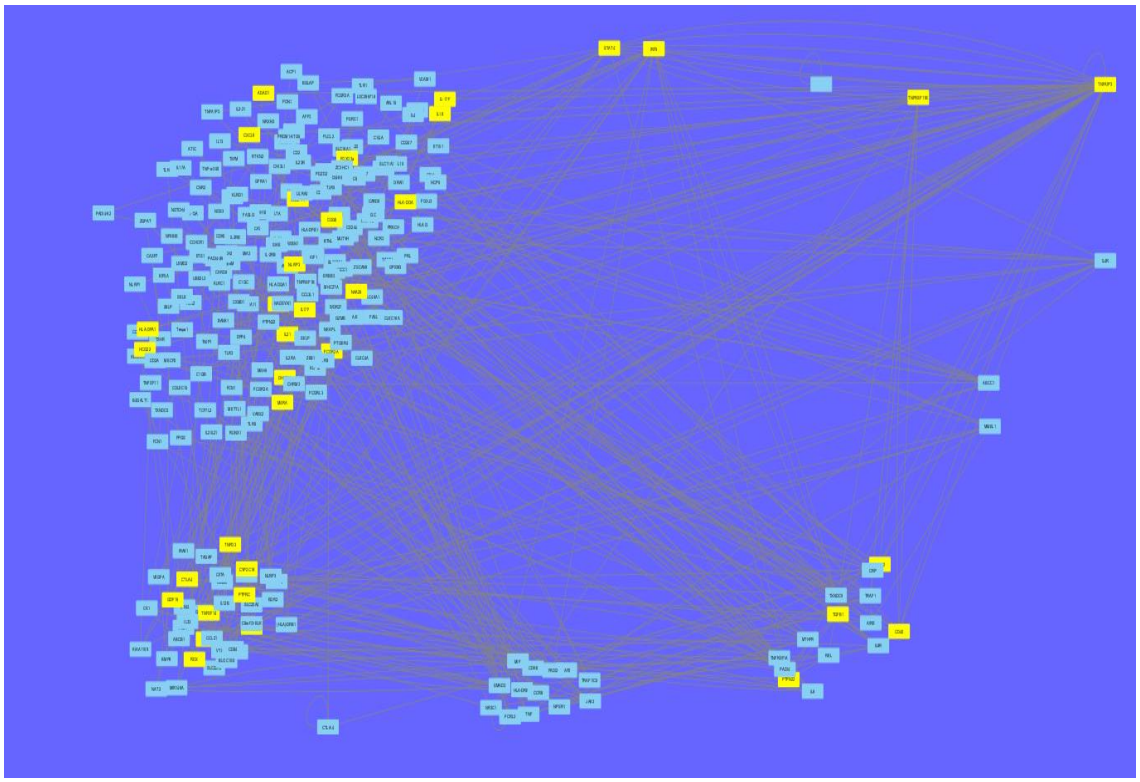


Figure 5.1Representation of connectivity between the nodes

In the current research we found that the gene TNFAIP3 with maximum degree which is 48 could act as a hub protein. The first neighbours of TNFAIP3 is TNFSF14,TNPO3,TGFB1, NLRP3 , NAA25, MSRA, PTPN2, IL21, RDX, IL18, IL17F, IRF5, ITGAV, HLA-DPA1, HLA, DOA, HCG22, IL17F, FCGR2A , GDF15, CYP2C19, CTLA4 , DHODH, COG6, CXCL9, CYB5A, STAT4, TNFRSF11B, CD40, CYP1A2, FOXO3a, PTPRC, PTPN22, ANGPT1, ADAD1.

In the Result panel of Cytoscape, The network statistics of Cytoscape includes various parameters as given below

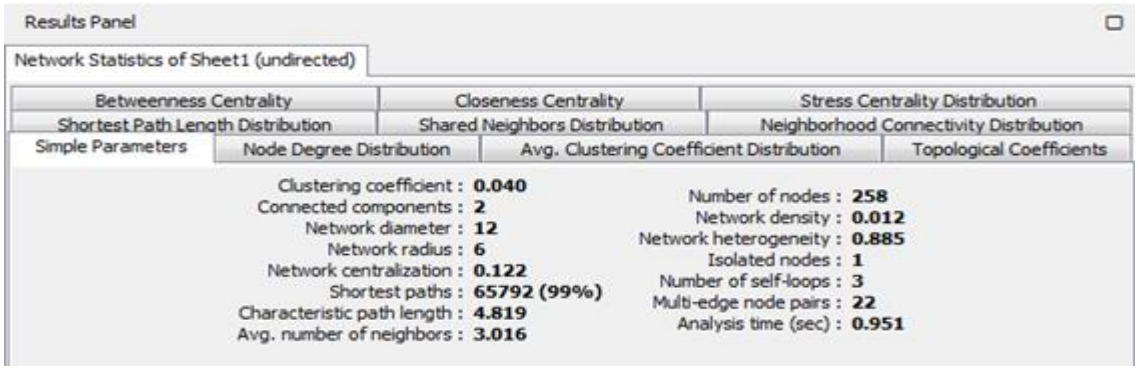


Figure 5.2 Result panel of the Network.

5.2.1 Betweeness centrality

Nodes with a high betweenness centrality is a measure of the number of shortest paths that go through each node. In networks with directed edges such as regulatory networks, bottlenecks tend to correlate with essentiality100. In the graph given below it has been shown that betweeness centrality ranges from.000 to 0.425.

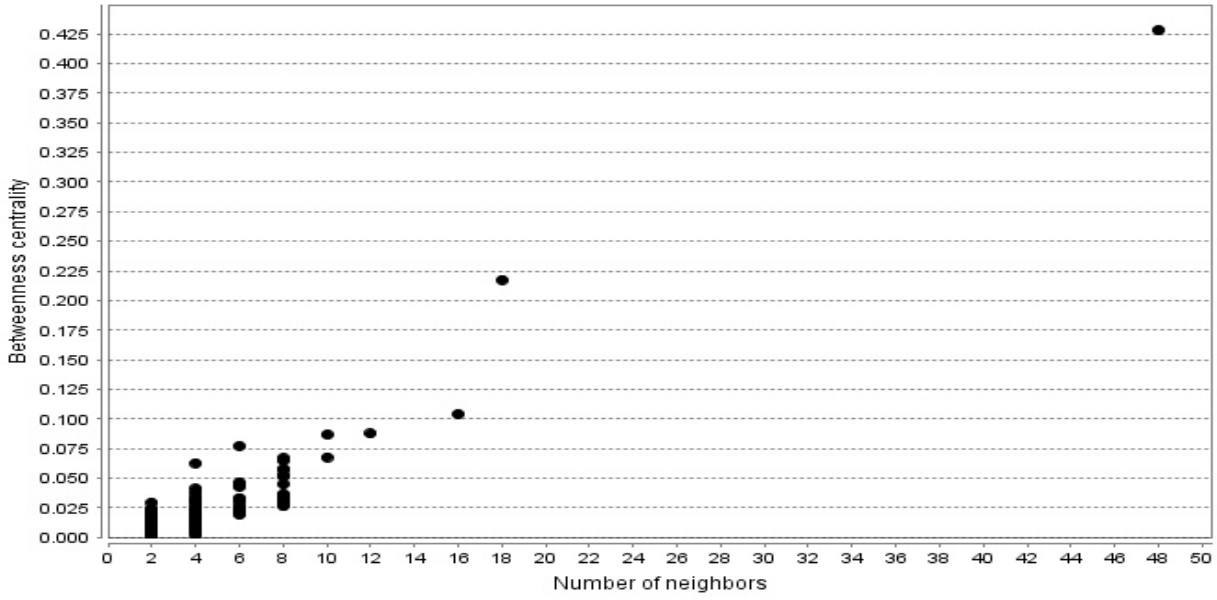


Figure5.3 betweenness centrality and number of neighbor.

5.2.2Shortest path length distribution

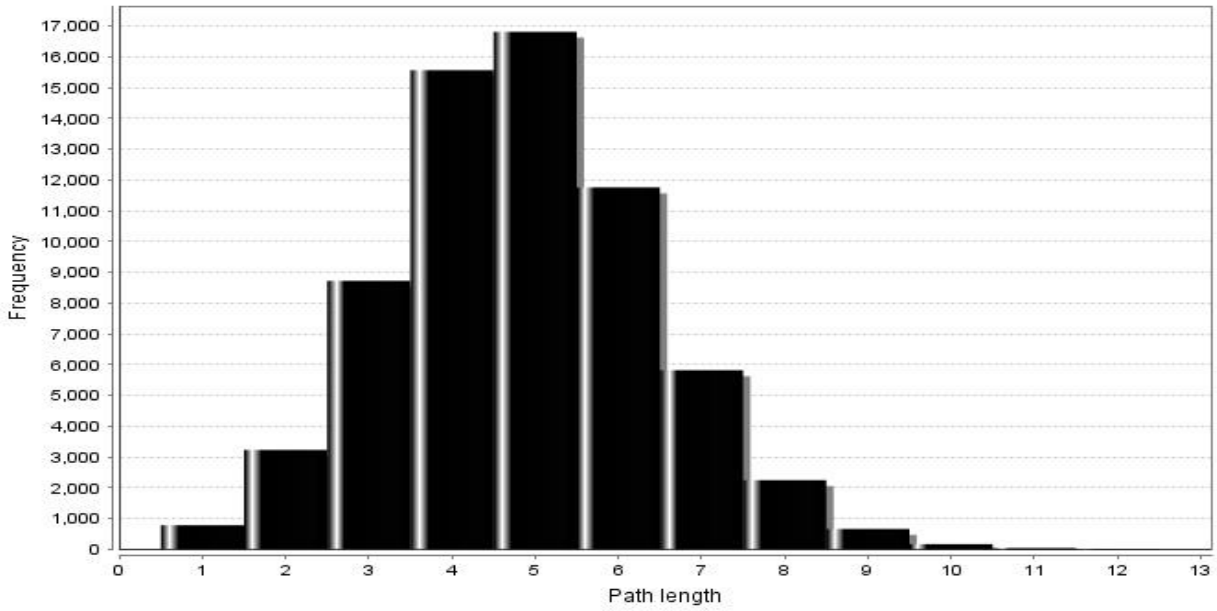


Figure 5.4 The calculation of shortest path length distribution using frequency and path length

Shortest path length facilitates the rapid transfer of information , in this graph we analyses the shortest path length of the network using comparative analysis between frequency and path length.

5.2.3Stress centrality

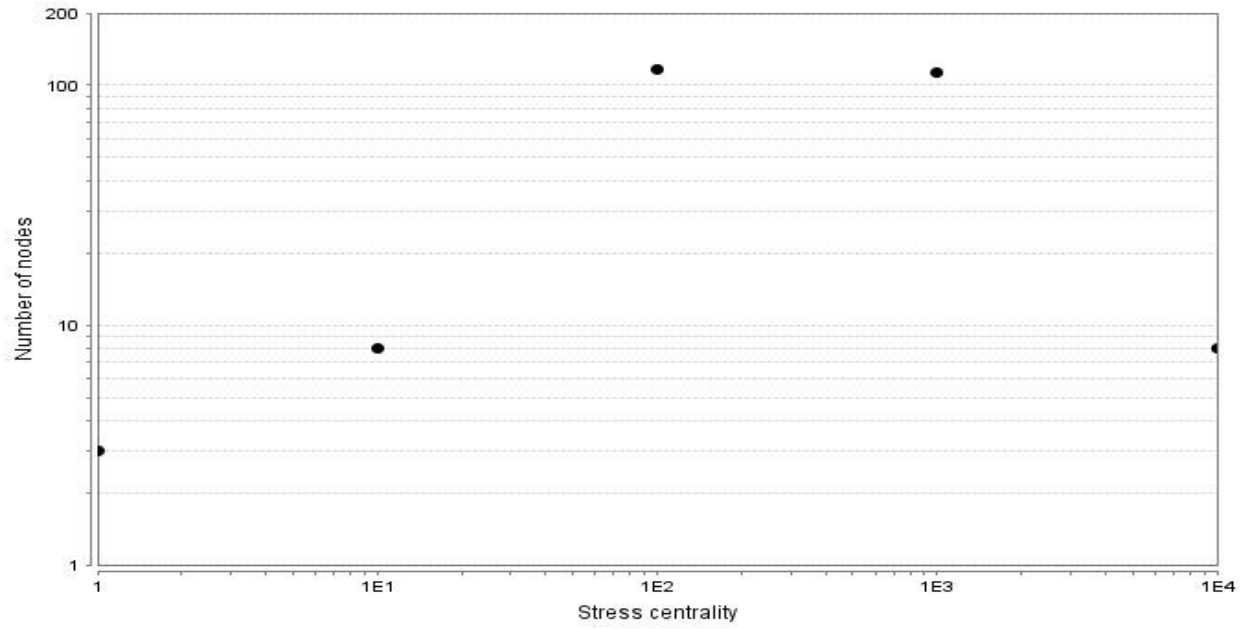


Figure 5.5 Relationship between stress centrality and number of nodes.

The betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes in the network. This measure favors nodes that join communities (dense subnetworks), rather than nodes that lie inside a community.

5.3 GENE ONTOLOGY ANALYSIS

The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects: **molecular function**(molecular activities of gene products) **cellular component**(where gene products are active) **biological process**(pathways and larger processes made up of the activities of multiple gene products).

GORILLA GENE ONTOLOGY

To test the Ontology of GOrilla we used the Rheumatoid Arthritis dataset, This dataset consists of 258 polymorphic genes obtained from literature annotation. Patients from different Geographical indication showed various characteristics.

5.3.1 Process analysis

The result of Process analysis is shown in figure 7 table 2 and highlights a unique set of enriched GO terms that were identified at different cutoffs.

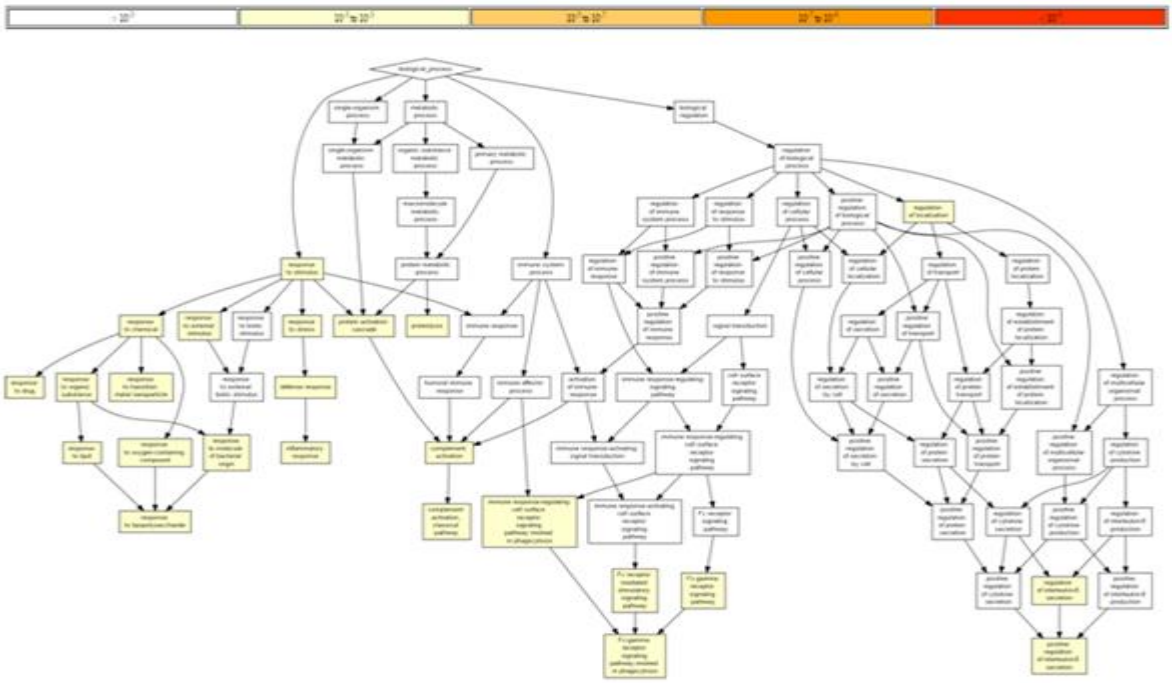


Figure 5.6 Process analysis of genes

The enriched GO terms include Response to chemical (top 46 genes), Response to external stimulus (top 49 genes) and Response to stimulus (top 85genes),Response to Stress (top 54 genes), Compliment activation, classical pathway (top 3 genes), Response to External biotic stimulus (top 42 genes), Response to biotic stimulus (top 42 genes), Response to organic substances (top 41 genes), Response to lipopolysaccharides (top 22 genes), Defence response (top 66 genes), Response to molecule of Bacterial origin (top 28 genes), Positive regulation of IL-8 secretion (top 3 genes),Response to Lipid(top 42 genes), Response to Oxygen containing compound(top 31 genes), Inflammatory response (44 genes), Regulation of Immune system process (top 53 genes), Positive regulation of immune system process (top 47 genes), Regulation of IL-8 secretion (top 5 genes), Regulation Of Localization(top 72 genes),Response to transition metal nanoparticle (top 2 genes), Fc receptor mediated stimulatory signalling pathway (top 3 genes), immune response regulatory cell surface receptor signalling pathway involved in Phagocytosis (top 3 genes) Fc gamma receptor signalling pathway involved in phagocytosis (top 3 genes), Fc gamma receptor signalling pathway involved in phagocytosis (top 3 genes), Fc gamma receptor signalling pathway (top

3 genes), Immune system process (top 85 genes), Proteolysis (top 4 genes), complement activation (top 3 genes), Protein activation cascade (top 3 genes).

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0042221	response to chemical	1.27E-5	5.07E-2	1.58 (211,74,83,46)	SELE,IL17A, LTA, CIITA,CXCL9,XRCC1, IL12A,MTHFR,PTGER4,SLC11A1,CD6,TLR1,CD58, NLRP3,DHODH,JAK3,TLR5, TLR4, ABCB1, CYP1A2, ETS1,CNR2, FOXO3, CHI3L1,SELP,PON1, BGLAP, VEGFA,IL4,PTPN22,C1QA,CCL3L1,CD9, ABCC1,IL10,CYB5A,TNF,TNFRSF1A,OGG18,TNFAIP3,VCAM1,TNFRSF1B,FASLG,TNFRSF11B,IL23R,IL6R
GO:0009605	response to external stimulus	2.01E-5	4.03E-2	1.48 (211,64,109,49)	IL17A,SELE,VTCN1,HLAE,CXCL9,IL12B,IL12A,MTHFR,SLC11A1,TLR1,TLR3,TLR5,TLR4,CHI3L1,SELP,IFNG,BGLAP,IL4, IL33,CARD9,TNF,IL10,TNFRSF1A,TNFAIP3,OGG1 – 8,VCAM1,TNFRSF1B,FASLG, IL6,TNFRSF11B,IL6R,IL23R,CRP,LTA,PTGER4,CD6,DHODH,,NLRP3,CYP1A2,ETS1,TYMS,CNR2,PON1,CDK6,PTPN22,FCN2,IRF5,SMAD3, PTPRC
GO:0050896	response to stimulus	3.03E-5	4.04E-2	1.25 (211,126,114,85)	IL17A,SELE,AIRE,VTCN1,CD244,CIITA, HLAE,CXCL9,XRCC1,IL12B,IL12A,MTHFR,DPP4,SLC11A1,CTLA4,TLR1,CD58, HLA-DQA1,TLR3, RDX, TLR5, ABCB1, TLR4,FOXO3,CHI3L1,IFNG,SELP,BGLAP,IL4R,VEGFA,IL4,PADI4,IL33,CD84,CCL3L1,CARD9,TNF,CYB5A,IL10,KLRD1,TNFRSF1A,CFH,TNFAIP3,OGG18,VCAM1, IL17F,B4GALT1,TNFRSF1B,FASLG,IL6,IL23R,TNFRSF11B,IL6R,CRP,LTA,PTGER4,CD6,TAP2,DHODH,NLRP3,JAK3,CYP1A2,ETS1,TYMS,CNR2,CDK6,PON1,PTPN22,FCN2,MUTYH,C1QA,C1QB,COLEC10,

					C1QC,FCGR3A,NR3C1,ABCC1,MIA3,ORAI1,IRF5,ERBB3,SMAD3,C5,PTPRC,MPG
GO:0006950	response to stress	4.4E-5	4.4E-2	1.44 (211,94,84,54)	IL17A,SELE,CD244, HLA,CXCL9,XRCC1,IL12A,MTHFR,SLC11A1,CTLA4,TLR1,TLR5,TLR4,FOXO3,CHI3L1,SELP,IL4,VEGFA,PADI4,IL33,CD84,CCL3L1,CARD9,IL10,TNF,TNFRSF1A,CFH,KLRD1,OGG1,TNFAIP3,VCAM1,B4GALT1,TNFRSF1B,IL23R,IL6R,TNFRSF11B,CRP,LTA,PTGER4,CD6,NLRP3,DHODH,JAK3,CYP1A2,ETS1,CNR2,FCN2,C1QA,MUTYH,C1QB,C1QC,MIA3,C5,PTPRC
GO:0006958	complement activation, classical pathway	6.64E-5	5.32E-2	31.65 (211,5,4,3)	C1QA,C1QB,C1QC
GO:0043207	response to external biotic stimulus	1.35E-4	9.03E-2	1.48 (211,53,113,42)	SELE,IL17A,CRP,VTCN1,LTA,HLAE,CL9,IL12B,IL12A,PTGER4,SLC11A1,CD6,TLR1,NLRP3,TLR3,TLR5,TLR4,CYP1A2,CNR2,SELP,IFNG,CDK6,IL4R,IL4,PTPN22,IL33,FCN2,CARD9,IL10,TNF,TNFRSF1A,TNFAIP3,IRF5,VCAM1,SMAD3,PTPRC,TNFRSF1B,FASLG,IL6,IL23R,IL6R, TNFRSF11B
GO:0009607	response to biotic stimulus	1.35E-4	7.74E-2	1.48 (211,53,113,42)	SELE,IL17A,CRP,VTCN1,LTA,HLAE,CXCL9,IL12B,IL12A,PTGER4,SLC11A1,CD6,TLR1,NLRP3,TLR3,TLR5,TLR4,CYP1A2,CNR2,IFNG,SELP,CDK6,IL4R,IL4,PTPN22,IL33,FCN2,CARD9,IL10,TNF,TNFRSF1A,TNFAIP3,IRF5,VCAM1,SMAD3,PTPRC,TNFRSF1B,FASLG,IL6,IL23R,IL6R, TNFRSF11B,
GO:0010033	response to organic substance	1.62E-4	8.11E-2	1.56 (211,67,83,41)	IL17A,SELE,LTA,CIITA,CXCL9,XRCC1,IL12A,MTHFR,PTGER4,SLC11A1,CD6,CD58,TLR1,NLRP3,DHODH,JAK3,TLR5,TLR4,CYP1A2,ETS1,CNR2,CHI3L1,SELP,PO

					N1,BGLAP,VEGFA,IL4,PTPN22, CCL3L1, CARD9,TNF,IL10,TNFRSF1A,OGG1,TNFAIP3,VCAM1,TNFRSF1B,FASLG,TNFRSF11B,IL23R, IL6R
GO:0032496	response to lipopolysaccharide	1.8E-4	8.01E-2	1.95 (211,29,82,22)	SELE,PTPN22,LTA,CXCL9,IL12A,SLC11A1,PTGER4,IL10,CD6,TNFRSF1A,NLRP3,TNFAIP3,VCAM1,TLR5,TLR4,CYP1A2,TNFRSF1B,FASLG,CNR2,IL23R,TNFRSF11B, SELP
GO:0006952	defense response	1.98E-4	7.91E-2	1.25 (211,75,148,66)	SELE,IL17A,IL18,CD244,HLA-E,CXCL9,IL12B,IL12A,DPP4,SLC11A1,TLR1,REL,TLR3,TLR5,TLR4,TLR8,CHI3L1,SELP,IFNG,CR1,IL4R,IL4,PADI4,TNIP1,IL33,IL2RA,CLEC6A,AIF1,CD40,CD84,CCL3L1,CARD9,IL37,IL10,TNF,KLRD1,CFH,TNFRSF1A,OGG1,TNFAIP3,B4GALT1,VCAM1,IL17F,TNFRSF1B,IL6,TNFRSF11B,IL6R,IL23R,CCR6,CRP,LTA,GZMB,CD6,TLR9,NLRP3,JAK3,CNR2,FCN2,C1QA,C1QB,C1QC,NCR3,IRF5,C5,PTPRC,IRAK1
GO:0002237	response to molecule of bacterial origin	2.37E-4	8.62E-2	1.64 (211,33,109,28)	SELE,LTA,CXCL9,IL12B,IL12A,SLC11A1,PTGER4,CD6,TLR1,NLRP3,TLR5,TLR4,CYP1A2,CNR2,SELP,IFNG,PTPN22,CARD9,IL10,TNFRSF1A,TNFAIP3,IRF5,VCAM1,TNFRSF1B,FASLG,IL6,TNFRSF11B,IL23R
GO:2000484	positive regulation of interleukin-8 secretion	2.63E-4	8.77E-2	19.78 (211,4,8,3)	CD58,CD244,CD2
GO:0033993	response to lipid	2.76E-4	8.51E-2	1.26 (211,42,168,42)	IL17A,SELE,LTA,IL13,CXCL9,IL12B,IL12A,PTGER4,SLC11A1,CD6,NLRP3,TLR5,TLR4,CYP1A2,ETS1,CNR2,TYMS,SELP,IFNG,PON1,BGLAP,IL4R,PTPN22,TGFB1,AIF1,CD40,NR3C1,TRAF6,IL10,TNF,TNFRSF1A,TNFAIP3,OGG1,VCAM1,TNFRSF

					1B,UBE2L3,FASLG,IL6,CCL21,TNFRSF11B,IRAK1,IL23R
GO:1901700	response to oxygen-containing compound	2.8E-4	8E-2	1.70 (211,47,82,31)	SELE,IL17A,LTA,CXCL9,IL12A,MTHFR,SLC11A1,PTGER4,CD6,NLRP3,TLR5,TLR4,CYP1A2,ETS1,CNR2,SELP,PON1,BGLAP,IL4,PTPN22,CARD9,TNF,IL10,TNFRSF1A,TNFAIP3,OGG18,VCAM1,TNFRSF1B,FASLG,TNFRSF11B,IL23R
GO:0006954	inflammatory response	2.9E-4	7.75E-2	1.24 (211,44,170,44)	IL17A,SELE,IL18,CRP,IL13,CXCL9,HLA-DRB1,SLC11A1,CD6,TLR1,REL,TLR9,NLRP3,TLR3,TLR5,TLR8,TLR4,CNR2,CHI3L1,SELP,TNIP1,TGFB1,IL2RA,AIF1,CD40,CCL3L1,NCR3,NLRP1,IL37,TNF,IL10,TNFRSF1A,OGG1,TNFAIP3,B4GALT1,VCAM1,IL17F,C5,TNFRSF1B,IL6,CCL21,IL6R,IL23R,TNFRSF11B
GO:0002682	regulation of immune system process	3.04E-4	7.61E-2	1.40 (211,94,85,53)	IL17A,VTCN1,CD244,HLAE,CXCL9,IL12A,SLC11A1,CTLA4,TLR1,TLR5,ABCB1,TLR4,FOXO3,SELP,BGLAP,IL4,VEGFA,IL33,CD84,CARD9,TNF,IL10,CFH,KLRD1,TNFAIP3,VCAM1,KLRC1,SIAE,IL6R,IL23R,LTA,CD226,PTGER4,CD6,NLRP3,JAK3,CD2,ETS1,CNR2,CD247,PTPN22,FCN2,TESPA1,C1QA,C1QB,COLEC10,C1QC,FCGR3A,FCGR2A,MIA3,C5,PTPRC,GPSM3
GO:0002684	positive regulation of immune system process	3.23E-4	7.6E-2	1.46 (211,80,85,47)	IL17A,VTCN1,CD244,LTA,HLAE,CXCL9,IL12A,CD226,PTGER4,SLC11A1,CD6,CTLA4,TLR1,NLRP3,JAK3,TLR5,ABCB1,TLR4,CD2,ETS1,FOXO3,CD247,SELP,VEGFA,IL4,PTPN22,IL33,FCN2,TESPA1,C1QA,C1QB,C1QC,COLEC10,CD84,FCGR3A,CARD9,FCGR2A,TNF,MIA3,CFH,TNFAIP3,VCAM1,C5,PTPRC,GPSM3, IL6R, IL23R,
GO:2000482	regulation of interleukin-8	4.51E-4	1E-1	6.51 (211,6,27,5)	CD58,CRP,CD244,PTPN22,CD2

	secretion				
GO:0032879	regulation of localization	5.51E-4	1.16E-1	1.18 (211,76,170,72)	SELE,CARD8,IL17A,IL18,VTCN1,CD244,HLA-E,CD2AP,CXCL9,IL13,IL12B,IL12A,HLADRB1,SLC11A1,TLR1,CD58,TLR3,RDX,TLR5,TLR8,TLR4,NOS3,BANK1,IFNG,SELP,IL4R,IL4,VEGFA,IL33,CLEC6A,AIF1,PADI2,CD84,CD40,NLRP1,NFKBIE,TNF,IL10,OGG1,NPSR1,IL17F,B4GALT1,CCL21,IL6,IL6R,PLB1,CCR6,CRP,GZMB,PTGER4,TLR9,NLRP3,JAK3,ITGAV,CD2,ETS1,HTR2A,PON1,CDK6,TGFB1,PTPN22,PTPN2,TRAF6,TCF7L2,MIA3,ORAI1,ERBB3,C5,SMAD3,PTPRC,GPSM3,UBE2L3
GO:1990267	response to transition metal nanoparticle	6.97E-4	1.4E-1	52.75 (211,4,2,2)	BGLAP C1QA
GO:0002431	Fc receptor mediated stimulatory signaling pathway	7.42E-4	1.41E-1	12.41 (211,3,17,3)	FCGR2A FCGR3A CD247
GO:0002433	immune response-regulating cell surface receptor signaling pathway involved in phagocytosis	7.42E-4	1.35E-1	12.41 (211,3,17,3)	FCGR2A FCGR3A CD247
GO:0038096	Fc-gamma receptor signaling pathway involved in phagocytosis	7.42E-4	1.29E-1	12.41 (211,3,17,3)	FCGR2A,CD247,FCGR3A
GO:0038094	Fc-gamma receptor signaling pathway	7.42E-4	1.24E-1	12.41 (211,3,17,3)	FCGR2A,FCGR3A,CD247

GO:0002376	immune system process	7.92E-4	1.27E-1	1.19 (211,104,145,85)	IL17A,SELE,AIRE,VTCN1,CD244,CIITA,HLAE,CXCL9,IL12B,IL12A,SLC11A1,DPP4,CTLA4,TLR1,CD58,REL,HLADQA1,TLR3,TLR5,ABCB1,TLR4,TLR8,BANK1,IFNG,SELP,CR1,IL4R,VEGFA,IL4,PADI4,TNIP1,IL33,IL2RA,CLEC6A,AIF1,CD84,CD40,CCL3L1,CARD9,IL37,TNF,IL10,TNFRSF1A,KLRD1,CFH,TNFAIP3,VCAM1,B4GALT1,TNFRSF1B,FASLG,IL6,IL6R,TNFRSF11B,CCR6,CRP,LTA,GZMB,PTGER4,CD6,TAP2,TLR9,NLRP3,JAK3,ITGAV,CD2,ETS1,CNR2,CD247,CDK6,PTPN22,FCN2,C1QA,C1QB,COLEC10,C1QC,FCGR3A,PTPN2,NCR3,FCGR2A,ORAI1,IRF5,SMAD3,C5,PTPRC,IRAK1
GO:0006508	Proteolysis	8.85E-4	1.36E-1	9.59 (211,22,4,4)	BGLAP,C1QA,C1QB, C1QC
GO:0006956	complement activation	9.84E-4	1.46E-1	17.58 (211,9,4,3)	C1QA,C1QB,C1QC
GO:0072376	protein activation cascade	9.84E-4	1.41E-1	17.58 (211,9,4,3)	C1QA,C1QB ,C1QC

Table 5.1 Process analysis of genes

5.3.2 Function analysis of genes

The result of Function analysis is shown in figure 5.7 table 5.2 and highlights a unique set of enriched GO terms that were identified at different cutoffs.

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0004252	serine-type endopeptidase activity	6.31E-4	4.53E-1	19.78 (211,8,4,3)	C1QA,C1QB,C1QC
GO:0008236	serine-type peptidase	6.31E-4	2.26E-1	19.78 (211,8,4,3)	C1QA,C1QB,C1QC

	activity				
GO:0017171	serine hydrolase activity	6.31E-4	1.51E-1	19.78 (211,8,4,3)	C1QA,C1QB,C1QC

Table 5.2 Function analysis of genes

The enriched GO terms include Serine-type endopeptidase activity (3 genes), Serine-type peptidase activity (*p* , 3 genes) and Serine hydrolase activity (*p* , 3 genes).

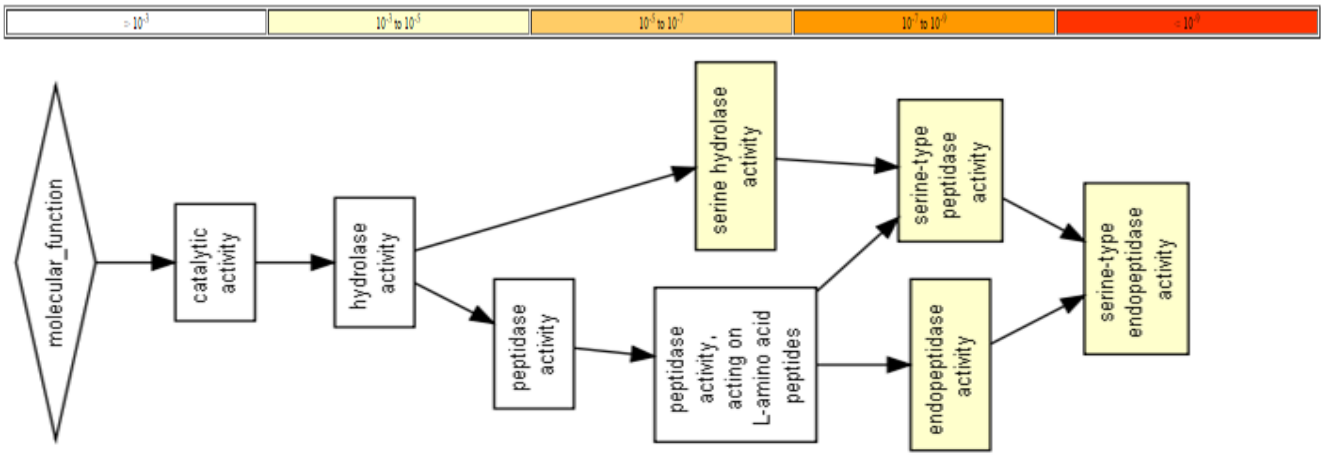


Figure 5.7 Function analysis of genes

5.3.3 Cellular component analysis

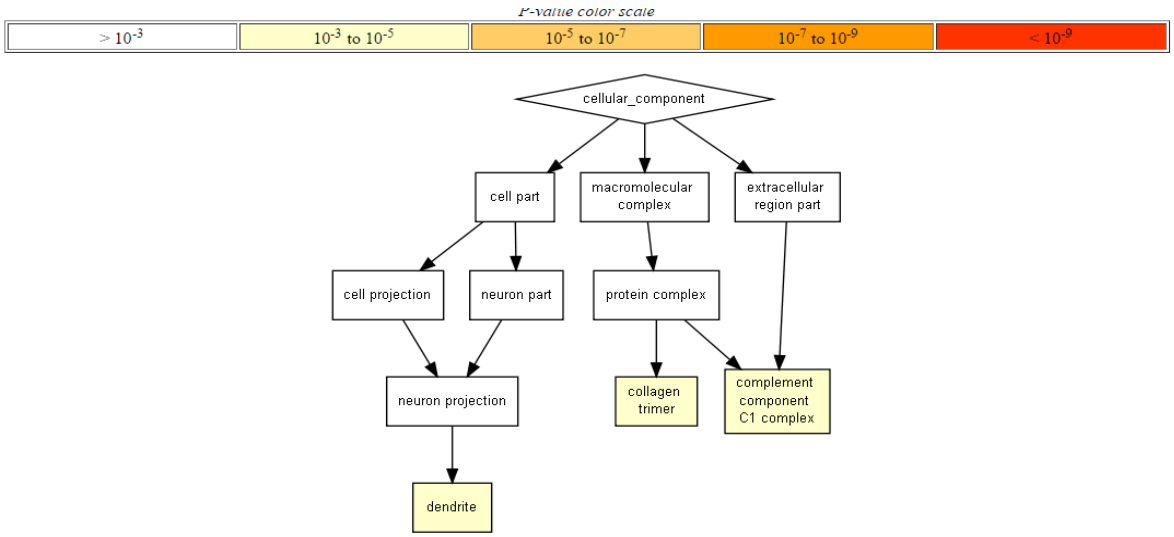


Figure 5.8 Cellular component analysis

The result of Function analysis is shown in figure 5.8 table 5.3 and highlights a unique set of enriched GO terms that were identified at different cutoffs. The enriched GO terms include Collagen trimmer (*p* , top 3 genes), compliment component C1 complex (*p* , top 2 genes) and Dendrite (*p* , top 3 genes).

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0005581	collagen trimer	1.64E-4	5.56E-2	26.38 (211,6,4,3)	C1QA,C1QB,C1QC
GO:0005602	complement component C1 complex	2.71E-4	4.6E-2	70.33 (211,2,3,2)	C1QA,C1QB
GO:0030425	Dendrite	3.14E-4	3.55E-2	16.23 (211,3,13,3)	BGLAP,CHRM3,CNR2

Table 5.3 Process analysis of genes

These enriched GO terms are attributed to genes that were over expressed in patients with bad prognosis and under-expressed in patients with good prognosis, which is in accordance with biological common sense and supports their relevance.

5.4 PPI ANALYSIS: STRING

In manual curation of data the validate interactions individually from the literature takes lots of efforts. Despite these curation efforts, the existing maps are taken as incomplete, and the literature-based datasets, although who have more interactions, are prone to investigative biases, having more interactions for the more explored ailment proteins.

The many function and interactions that occur between proteins are at the core of cellular processing and their systematic description helps to provide context in molecular systems biology. However, known and predicted interactions are spread all over multiple resources, and the available data exhibit remarkable differences in terms of quality and completeness. The STRING database intends to provide a critical assessment and integration of protein–protein interactions, including direct as well as indirect links. For this purpose, there are hierarchical and self-consistent orthology annotations for all interacting proteins at various levels of phylogenetic resolution.

There are functional interdependencies between the molecular components in a cell, a disease is not often a consequence of an abnormality in a single gene, but reflects the perturbations of the complex cellular network that links tissue and organ systems. These expanding tools of network offer a platform to explore systematically not only the molecular complexity of a particular ailment, leading to the recognition of disease modules and pathways, but also the molecular relationships among distinct phenotypes. Advancement in this direction are essential for classify new disease genes, for uncovering the biological importance of disease-associated mutations identified by GWAS and full-genome sequencing, and for identifying drug targets and biomarkers for complex ailments.

In PPI analysis we mainly analyses the Hub protein interaction with seed genes. TNFAIP3 shows the direct neighbouring with TNIP1, BANK1, WDFY4, PTPN22, IL17A, STAT4, TNF, LTA, NFKBIE, TNFRSF1A, REL,TRAF6, CD40, IRF5, TRAF1.

If Hub gene TNFAIP3 analyses separately than TNFAIP3 shows interaction with TNIP2, TNIP1, TAX1BP1, TBK1, IKBKG, TRAF6,TRAF2, RIPK1, IKBKB, UBC. The genes TNIP1, TRAF6 common in both with seed genes and separate interactions.

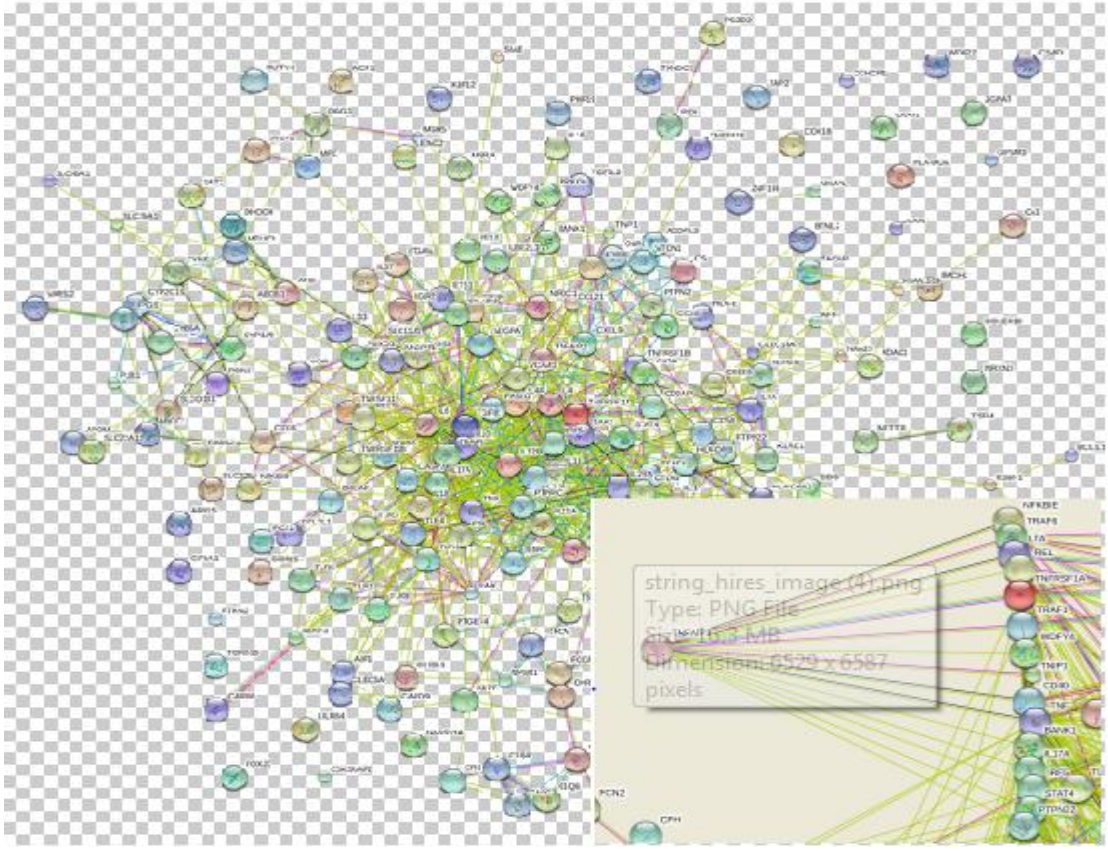


Figure 5.9 PPI analysis

5.5.1 DAVID Functional Annotation Chart

Functional Annotation Chart provides typical gene-term enrichment analysis, that is also provided by other similar tools, to identify the most relevant biological terms associated with a seed genes .Compared to other similar enrichment analysis tools, the notable difference of this function provided by DAVID is its extended annotation coverage, increasing from only GO in the original version of DAVID to currently over 40 annotation categories, including GO terms, protein–protein interactions, protein functional domains, disease associations, bio-pathways, sequence features, homology, gene functional summaries, gene tissue expression, and literature.

These are the enriched Pathways in Functional analysis on the basis of filtered p value less than 0.01 as, Allograft rejection, Cytokine-cytokine receptor interaction, Graft-versus-host disease, Type 1 diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, cell adhesion molecules, Toll like receptor signalling pathway, intestinal immune network for IgA production, JAK-STAT signalling pathway, Antigen processing and presentation, NOD- like receptor signalling pathway,natural killer cell mediated cytotoxicity, hematopoietic cell lineage.

	Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enric	Bonferror	Benjamin	FDR
2	KEGG_PA	hsa05330:Allograft rejection	14	0.60896	5.96E-13	804491, 79	122	36	5085	16.20902	6.26E-11	6.26E-11	6.67E-10
3	KEGG_PA	hsa04060:Cytokine-cytokine receptor interaction	30	1.304915	1.04E-12	818161, 79	122	262	5085	4.772557	1.10E-10	5.48E-11	1.17E-09
4	KEGG_PA	hsa05332:Graft-versus-host disease	13	0.565463	4.32E-11	778475, 78	122	39	5085	13.89344	4.54E-09	1.51E-09	4.84E-08
5	KEGG_PA	hsa04940:Type I diabetes mellitus	12	0.521966	2.01E-09	778475, 78	122	42	5085	11.90867	2.11E-07	5.28E-08	2.25E-06
6	KEGG_PA	hsa05310:Asthma	10	0.434972	1.22E-08	778475, 78	122	29	5085	14.37253	1.28E-06	2.56E-07	1.36E-05
7	KEGG_PA	hsa05320:Autoimmune thyroid disease	12	0.521966	1.87E-08	778475, 78	122	51	5085	9.807136	1.96E-06	3.27E-07	2.09E-05
8	KEGG_PA	hsa05322:Systemic lupus erythematosus	15	0.652458	6.05E-08	796581, 80	122	99	5085	6.315201	6.36E-06	9.08E-07	6.78E-05
9	KEGG_PA	hsa04514:Cell adhesion molecules (CAMs)	17	0.739452	6.07E-08	787386, 80	122	132	5085	5.367921	6.37E-06	7.96E-07	6.79E-05
10	KEGG_PA	hsa04620:Toll-like receptor signaling pathway	14	0.60896	5.72E-07	804552, 80	122	101	5085	5.777471	6.00E-05	6.67E-06	6.40E-04
11	KEGG_PA	hsa04672:Intestinal immune network for IgA production	10	0.434972	1.67E-06	778475, 78	122	49	5085	8.506189	1.76E-04	1.76E-05	0.001875
12	KEGG_PA	hsa04630:Jak-STAT signaling pathway	15	0.652458	1.50E-05	804081, 79	122	155	5085	4.03358	0.001572	1.43E-04	0.016782
13	KEGG_PA	hsa04612:Antigen processing and presentation	11	0.478469	2.25E-05	778475, 78	122	83	5085	5.523899	0.00236	1.97E-04	0.025204
14	KEGG_PA	hsa04621:NOD-like receptor signaling pathway	9	0.391475	9.40E-05	825577, 77	122	62	5085	6.05037	0.009826	7.59E-04	0.105286
15	KEGG_PA	hsa04650:Natural killer cell mediated cytotoxicity	12	0.521966	2.85E-04	802720, 81	122	133	5085	3.760631	0.029475	0.002135	0.318656
16	KEGG_PA	hsa04640:Hematopoietic cell lineage	9	0.391475	9.17E-04	778475, 82	122	86	5085	4.361895	0.091816	0.0064	1.022145

Table 5.4 Kegg Pathway analysis

The annotation categories can be flexibly included or excluded from the analysis. The enhanced annotation coverage alone increases the analytic power to analyze the genes from many different biological aspects in a single space. In addition, to take full advantage of the well-known KEGG and BioCarta pathways, DAVID Pathway Viewer, which is accessed by

clicking on pathway links within the chart report, can display genes from a user's list on pathway maps to facilitate biological interpretation in a network context. Finally, the choice of pre-built or user-defined gene population backgrounds provides the ability of tailor the enrichment analysis to meet the specific analytic situation.

5.5.2 DAVID Functional Annotation Clustering

Functional Annotation Clustering uses a similar fuzzy clustering concept as Functional Classification by measuring relationships among the annotation terms based on the degree of their co-association with genes in order to cluster somewhat heterogeneous, yet highly similar annotation into functional annotation groups. This reduces the burden of associating different terms associated with the similar biological process, thus allowing the biological interpretation to be more focused at the "biological module" level. The 2-D view tool is also provided for examining the internal relationships among the clustered terms and genes. This type of grouping of functional annotation is able to give a more insightful view of the relationships between annotation categories and terms compared to the traditional linear list of enriched terms since highly related/redundant annotation terms may be dispersed among hundreds, if not thousands, of other terms.

These pathways are expressed in enriched cluster on the basis of p value less than .001 Allograft rejection, graft verses host disease, Type 1 diabetes mellitus, Asthma, Autoimmune thyroid disease, Systemic lupus erythematosus, intestinal immune network for IgA production, Antigen processing and presentation.

1	Annotatio	Enrichment Score: 7.417379834795882											
2	Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enric	Bonferron	Benjamini	FDR
3	KEGG_PA	hsa05330:Allograft rejection	14	0.60896	5.96E-13	804491, 7	122	36	5085	16.209	6.26E-11	6.26E-11	6.67E-10
4	KEGG_PA	hsa05332:Graft-versus-host disease	13	0.56546	4.32E-11	778475, 7	122	39	5085	13.8934	4.54E-09	1.51E-09	4.84E-08
5	KEGG_PA	hsa04940:Type I diabetes mellitus	12	0.52197	2.01E-09	778475, 7	122	42	5085	11.9087	2.11E-07	5.28E-08	2.25E-06
6	KEGG_PA	hsa05310:Asthma	10	0.43497	1.22E-08	778475, 7	122	29	5085	14.3725	1.28E-06	2.56E-07	1.36E-05
7	KEGG_PA	hsa05320:Autoimmune thyroid disease	12	0.52197	1.87E-08	778475, 7	122	51	5085	9.80714	1.96E-06	3.27E-07	2.09E-05
8	KEGG_PA	hsa05322:Systemic lupus erythematosus	15	0.65246	6.05E-08	796581, 8	122	99	5085	6.3152	6.36E-06	9.08E-07	6.78E-05
9	KEGG_PA	hsa04672:Intestinal immune network for IgA production	10	0.43497	1.67E-06	778475, 7	122	49	5085	8.50619	1.76E-04	1.76E-05	0.00187
10	KEGG_PA	hsa04612:Antigen processing and presentation	11	0.47847	2.25E-05	778475, 7	122	83	5085	5.5239	0.00236	1.97E-04	0.0252

Table 5.5 Function analysis

Highest kappa score showing by these pathways as

	Category	Term	Kappa
1	KEGG_PAT	hsa05330:Allograft rejection	1
2	KEGG_PAT	hsa05320:Autoimmune thyroid disease	0.845284
3	KEGG_PAT	hsa04940:Type I diabetes mellitus	0.845284

Table 5.6 highly similar cluster pathways

CHAPTER 6

CONCLUSIONS

We find that the new method makes a considerable difference to the categories identified as the most significant. Here we have developed a gene interaction network analysis for use with polymorphic genes of Rheumatoid Arthritis. This analysis is done using Literature survey and we consider only polymorphic genes which are validated by PUBMED literature. Literature survey provides a list of seed genes which provide hub genes through network analysis, important pathways through ontology analysis, functional enrichment through DAVID etc. Furthermore, we find that the most significant categories identified using network analysis, GO analysis, functional enrichment analysis and try to find out the better analysis then existing analysis of Rheumatoid Arthritis.

Areas of related and further research

The disease module which we obtained for Rheumatoid Arthritis further use in Pharmaco-genomics study and treatment of Rheumatoid Arthritis.

REFERENCES

Ahmadloo, Somayeh, et al. "Single Nucleotide Polymorphism rs 2476601 of PTPN22 Gene and Susceptibility to Rheumatoid Arthritis in Iranian Population." *Iranian Journal of Allergy, Asthma and Immunology* 14.4 (2015): 437.

Bakir-Gungor, Burcu, and Osman Ugur Sezerman. "A new methodology to associate SNPs with human diseases according to their pathway related context." *PloS one* 6.10 (2011): e26277.

Barabasi, A.L., Gulbahce, N. and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. Nat. Rev. Genet., 12, 56–68.

Bindea, Gabriela, et al. "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks."*Bioinformatics* 25.8 (2009): 1091-1093.

Camargo, Anyela, and Francisco Azuaje. "Identification of dilated cardiomyopathy signature genes through gene expression and network data integration." *Genomics* 92.6 (2008): 404-413.

Camargo, Anyela, and Francisco Azuaje. "Linking gene expression and functional network data in human heart failure." *PLoS One* 2.12 (2007): e1347.

Carrara, Greta, et al. "A validation study of a new classification algorithm to identify rheumatoid arthritis using administrative health databases: case–control and cohort diagnostic accuracy studies. Results from the RECord linkage On Rheumatic Diseases study of the Italian Society for Rheumatology." *BMJ open* 5.1 (2015): e006029.

Choy, Ernest. "Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis." *Rheumatology* 51.suppl 5 (2012): v3-v11.

Clément-Ziza, Mathieu, et al. "Genoscape: a Cytoscape plug-in to automate the retrieval and integration of gene expression data and molecular networks." *Bioinformatics* 25.19 (2009): 2617-2618.

Cline, Melissa S., et al. "Integration of biological networks and gene expression data using Cytoscape." *Nature protocols* 2.10 (2007): 2366-2382.

Deshmukh, Harshal A., et al. "Evaluation of 19 autoimmune disease-associated loci with rheumatoid arthritis in a Colombian population: evidence for replication and gene-gene interaction." *The Journal of rheumatology* 38.9 (2011): 1866-1870.

Dickerson, Jonathan E., et al. "Defining the role of essential genes in human disease." *PloS one* 6.11 (2011): e27368.

Doyle, Joseph P., et al. "Application of a translational profiling approach for the comparative analysis of CNS cell types." *Cell* 135.4 (2008): 749-762.

Eden, Eran, et al. "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists." *BMC bioinformatics* 10.1 (2009): 1.

Figeys, Daniel. "Mapping the human protein interactome." *Cell research* 18.7 (2008): 716-724.

Freudenberg, Jan, Peter Gregersen, and Wentian Li. "Enrichment of genetic variants for rheumatoid arthritis within T-cell and NK-cell enhancer regions."*Molecular Medicine* 21.1 (2015): 180.

Fronczuk, Maciej, Adrian E. Raftery, and Ka Yee Yeung. "CyNetworkBMA: a Cytoscape app for inferring gene regulatory networks." *Source code for biology and medicine* 10.1 (2015): 1.

Holmans, Peter, et al. "Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder." *The American Journal of Human Genetics* 85.1 (2009): 13-24.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." *Nucleic acids research* 37.1 (2009): 1-13.

Huang, Da Wei, et al. "The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists." *Genome biology* 8.9 (2007): 1.

Hunt, Arthur G., et al. "Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling." *BMC genomics* 9.1 (2008): 1.

Kanehisa, Minoru, and Peer Bork. "Bioinformatics in the post-sequence era."*nature genetics* 33 (2003): 305-310.

Kelly, Victoria, and Mark Genovese. "Novel small molecule therapeutics in rheumatoid arthritis." *Rheumatology* (2013): kes367.

Lai, Eric. "Application of SNP technologies in medicine: lessons learned and future challenges." *Genome Research* 11.6 (2001): 927-929.

Lopes, Christian T., et al. "Cytoscape Web: an interactive web-based network browser." *Bioinformatics* 26.18 (2010): 2347-2348.

Maere, Steven, Karel Heymans, and Martin Kuiper. "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks." *Bioinformatics* 21.16 (2005): 3448-3449.

Malik, Rainer, et al. "From proteome lists to biological impact–tools and strategies for the analysis of large MS data sets." *Proteomics* 10.6 (2010): 1270-1283.

Meyer, Joel N., et al. "Decline of nucleotide excision repair capacity in aging Caenorhabditis elegans." *Genome biology* 8.5 (2007): 1.

Nagaraj, Shivashankar H., and Antonio Reverter. "A Boolean-based systems biology approach to predict novel genes associated with cancer: Application to colorectal cancer." *BMC systems biology* 5.1 (2011): 1.

Nakai, Yuji, et al. "Up-regulation of genes related to the ubiquitin-proteasome system in the brown adipose tissue of 24-h-fasted rats." *Bioscience, biotechnology, and biochemistry* 72.1 (2008): 139-148.

Natale, Massimo, et al. "FunMod: A Cytoscape Plugin for Identifying Functional Modules in Undirected Protein–Protein Networks." *Genomics, proteomics & bioinformatics* 12.4 (2014): 178-186.

Nayak, Renuka R., et al. "Coexpression network based on natural variation in human gene expression reveals gene interactions and functions." *Genome research* 19.11 (2009): 1953-1962.

Ogdie, Alexis, et al. "Risk of major cardiovascular events in patients with psoriatic arthritis, psoriasis and rheumatoid arthritis: a population-based cohort study." *Annals of the rheumatic diseases* (2014): annrheumdis-2014.

Östlund, Gabriel, Mats Lindskog, and Erik LL Sonnhammer. "Network-based Identification of novel cancer genes." *Molecular & Cellular Proteomics* 9.4 (2010): 648-655.

Pavlopoulos, Georgios A., et al. "Medusa: A tool for exploring and clustering biological networks." *BMC research notes* 4.1 (2011): 1.

Praneenararat, Thanet, Toshihisa Takagi, and Wataru Iwasaki. "Integration of interactive, multi-scale network navigation approach with Cytoscape for functional genomics in the big data era." *BMC genomics* 13.7 (2012): 1.

Shannon, Paul, et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." *Genome research* 13.11 (2003): 2498-2504.

Sharma, Amitabh, et al. "A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes." *Human molecular genetics* (2015): ddv001.

Sharma, Amitabh, et al. "Gene prioritization in Type 2 Diabetes using domain interactions and network analysis." *BMC genomics* 11.1 (2010): 1.

Solinas, Antonio, et al. "Duplex Scorpion primers in SNP analysis and FRET applications." *Nucleic Acids Research* 29.20 (2001): e96-e96.

Stephens, Susie, et al. "Aggregation of bioinformatics data using Semantic Web technology." *Web Semantics: Science, Services and Agents on the World Wide Web* 4.3 (2006): 216-221.

Stevens, Adam, et al. "Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks." *BMC genomics* 14.1 (2013): 1.

Su, Gang, et al. "GLay: community structure analysis of biological networks." *Bioinformatics* 26.24 (2010): 3135-3137.

Taylor, R. James, Andrew F. Siegel, and Timothy Galitski. "Network motif analysis of a multi-mode genetic-interaction network." *Genome biology* 8.8 (2007): 1.

Thomas, Paul D., et al. "Applications for protein sequence–function evolution data: mRNA/protein expression analysis and coding SNP scoring tools."*Nucleic acids research* 34.suppl 2 (2006): W645-W650.

Thomas, Sterling, and Danail Bonchev. "A survey of current software for network analysis in molecular biology." *Human genomics* 4.5 (2010): 1.

Tian, Chao, et al. "Analysis and application of European genetic substructure using 300 K SNP information." *PLoS Genet* 4.1 (2008): e4.

Van Steenbergen, Hanna W., et al. "A genetic study on C5-TRAF1 and progression of joint damage in rheumatoid arthritis." *Arthritis research & therapy* 17.1 (2015): 1.

Vlasblom, James, et al. "GenePro: a Cytoscape plug-in for advanced visualization and analysis of interaction networks." *Bioinformatics* 22.17 (2006): 2178-2179.

Yang, Tsun-Po, et al. "Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies."*Bioinformatics* 26.19 (2010): 2474-2476.

Young, Matthew D., et al. "Gene ontology analysis for RNA-seq: accounting for selection bias." *Genome biology* 11.2 (2010): 1.