Study of interactome for determining novel pathways and genes in vitiligo

To be submitted as major project report in partial fulfillment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Mayank kumar Singh

(DTU/14/M.Tech/119)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. YASHA HASIJA
Assistant Professor
Dept. Of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering, University of Delhi)

CERTIFICATE

This is to certify that the dissertation entitled **"*Study of interactome for determining novel pathways and genes in vitiligo*"** 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

DECLARATION

I declare that my project entitled "***Study of interactome for determining novel pathway and genes in vitiligo***" 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 a major project.

Date

Mayank Kumar Singh

2k14/bio/10

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisor and mentor, ***Dr. Yasha Hasija*** *who has guided me through thick and thin. without her valuable guidance , incessant help, calm endurance, constructive criticism, and constant encouragement the completion of this minor project would have been impossible the congenial discussion with her made me confident and gave enough independence in performing the experiment which enabled in me a higher efficiency to complete the work.*

Mayank kumar singh
2K14/bio/10

CONTENTS

List of figures:

List of tables

Title - ***Study of interactome for determining novel pathway and genes in vitiligo***

Name- Mayank kumar singh

Roll no.- 2k14/bio/10

Email id- mayank.singh91088@gmail.com

ABSTRACT

The approach rely on the assumption, assuming that if a few disease machineries are identified ,other disease-related components are likely to be found in their network vicinity. Therefore, The vitiligo module also contains immune response mechanisms that are common with other immune-related disease modules. Further, using diverse omics (genomics, gene-expression) data, we are identifying important signaling pathway as a novel modulator in vitiligo. Given the many SNPs , genes and environmental factors associated to the disease, traditional single gene or single pathway based approaches have shown limited utility. In the past not many years, several attempts have been made to incorporate the topological properties of protein interaction networks with dissimilar types of 'omics' data to discover novel genes and pathway search. disease can be associated to a well defined local neighborhood of the interactome. Yet, the existence of such a single vicinity remains a hypothesis that needs to be tested. During this work lots of knowledge about the network and its interaction is obtained . Using these result comparative and comphrensive analysis between nodes and its interacting neighborhood can be obtained . study of interactome provide the knowledge about the interaction between the genes and provide a clue about the factors responsible for vitiligo .These factors might include protein , hormones, enzyme or some other factors may also include . It also provide knowledge on the genes and how it is related to vitiligo and other diseases . Comphresnsive knowledge about the pathways involved in vitiligo is also found out. This is a very genuine and authentic study on vitiligo which could provide knowledge on pathway , genes , and factors involved in vitiligo and its process of spreading. Very interesting results and comphrensive analysis were found out during the whole study which could provide vital clue in the coming time related to vitiligo.

REVIEW OF LITERATURE

Vitiligo a type of skin disease found all over the world . In this work the target is to found out the genes and the factor responsible for this disease . vitiligo is very common in African and asian countries. Lots of research work has been done for finding out the factors responsible for the disease yet lot to be done as there is a limited research work that has been done in this area for finding out factors or genes or protein causing this disease. So to get true and validated result lots of bioinformatics tools are used such as DAVID bioinformatics tool , Pubmed , NCBI , Gorilla gene ontology , Cytoscape , GWAS , Dbsnp , String data base , KEGG pathway and many more. Lots of research papers and journals were looked upon by to get genuine , authentic and validated result . Using these tools many validated result , some complementing while other contradicting each other . Work on this disease is limited so to gather information related to vitiligo is quite difficult task because genes and proteins involved in this disease is very few. Network formation and its validation is very important so to validate result DAVID bioinformatics tool is used. In this work there is no wet lab work is performed and the result obtained is solely based on computational understanding and knowledge. On using all the bioinformatics tool which could be useful in analyzing and understanding this particular disease lots of interesting result is obtained. These result are based on some conclusions like :

-No. of chromosome involved in this particular disease.

-No of genes involved in vitiligo

-Factors such as protein or snps which are responsible for this particular disease.

-Pathways through which it get spread

-Factors of this disease common to two or more disease.

Apart from this many more interesting and concluding result are obtained . The main objective during the project work is to seek knowledge about the genes and many factors involved in vitiligo and finding out how these genes are influencing other genes and give rise to disease symptoms . Interctome study and pathway analysis during the work provide vital and important knowledge to give the answers to many questions raised in relation with vitiligo.

INTRODUCTION

Introduction to vitiligo

Vitiligo is a persistent skin condition characterized by some part of the skin loosing their pigment. It occurs when part of skin pigment cells get die or are unable to function. Apart from cases of contact with certain chemicals the reason for the cause of vitiligo is unknown. Research suggests that vitiligo may arise from autoimmune, oxidative stress , genetic, neural, or viral causes

Fig 1: Whitening of skin which is the major symptom of the vitiligo.

Symptoms :

The only mark of vitiligo is the occurrence of pale patchy areas of depigmented skin which tend to occur on the extremity. The patches are initially very small, but often grow and change shape and space. When skin lesion occur, they are mostly prominent on the face, hands and wrists. The loss of skin pigment is particularly visible around body orifices, such as the mouth, eyes, nostrils and genitalia . Some lesions have increased skin pigmentation around the edges.

Causes :

There are many hypothesis that have been proposed from time to time as potential reason for vitiligo but studies strongly suggests that change in the immune system is responsible for the condition related to vitiligo. It is found that vitiligo is a multifactorial disease with many genetic susceptibility and environmental factors both thought to play a role in increasing the disease

symptoms. The TYR gene which codes for the protein tyrosinase, which is not a component of the immune system,or related to that but it is an enzyme of the melanocyte that catalyze melanin biosynthesis, and a major autoantigen in vitiligo. In Some states the sunburns can cause the disease but there is not good evidence to support this hypothesis that there is any role of sunburn in causing vitilgo.

Genetic basis of vitiligo

As it is earlier mentioned that vitiligo is generally a disease related to depigmentation of the skin and is common in 0.2 % population of the world there is a limited number of studies were performed in the past to find out factors responsible for this particular disease. There were many hypothesis and concepts were given from time to time to describe the machinery involved in causing this disease. It can be explained by simple mendelian genetics and is characterized by incomplete penetrance, multiple susceptibility loci and genetic heterogeneity. NO doubt that environment play a significant role in propagation of vitiligo but the major cause for this disease is alteration in the genes related to vitiligo . Through various studies it is very clear that these influences are very complex and difficult to understand. Many studies have also provided strong support for vitiligo susceptibility genes on chromosomes 4q13–q21, 1p31, 7q22, 8p12 and 17p13, while loci of interest at 6p, 6q, 14q, 9q, 13q, 19p and 22q. This is mainly an autoimmune disease . Two main genes involved in vitiligo are *NLRP1* and *PTPN22.NLRP1* gene is mainly responsible for providing instruction for making protein which is involved in immune system which is responsible for the process of inflammation. Inflammation occurs when the immune system sends signaling molecules and white blood cells to a site of injury or disease to fight microbial invaders and facilitate tissue repair. The body then stops (inhibits) the inflammatory response to prevent damage to its own cells and tissues. The *PTPN22* gene provides instructions for making a protein involved in signaling that helps control the activity of immune system cells called T cells. T cells identify foreign substances and defend the body against infection. The variations in the *NLRP1* and *PTPN22* genes that are associated with an increased risk of developing vitiligo likely affect the activity of the *NLRP1* and *PTPN22* proteins, making it more difficult for the body to control inflammation and prevent the immune system from attacking its own tissues.

Protein Protein interaction network

Modelling and analyzing protein-protein interaction (PPI) networks is an important in systems biology. Many random graph models were found out to capture specific network characteristic properties or mimic the way genuine PPI networks might have evolved. Thus, our new approach allows us to look for high quality parts of currently available PPI data to create accurate models for PPI networks of different species. So in the process of data collection lots of journals and papers were searched to get accurate data. Most of the cellular functions are not performed out by single proteins, but by proteins acting together in a group. Due to the recent advancement in the experimental biological techniques such as tandem affinity purification(tap) and other high-throughput methods, there are giant amount of protein-protein interaction data publicly available which can be used according to our need. This large amount of data requires new mathematical and computational approache to be developed to analyze the complex network that is formed using theses mathematical and computational approach. An accurate model of PPI networks will allow to get better estimate of all types of network statistics for generating artificial networks of species for which protein-protein interactions have not been found out experimentally validated or determined. This could help understand and find out cellular processes and lead future biological experiments from which lots of information can be obtained Therefore, analyzing and modeling PPI networks has become a vast area of research . A PPI network is a graph formed with nodes (vertices) corresponding to proteins and edges (links) corresponds to relationship between the proteins.

Learning Network Structure

To learn the distribution we need to have a data set of 3-dimensional coordinates of "n" nodes in the Euclidean unit cube. These coordinates are the output of the algorithm applied to the PPI network data that we want to use for learning purpose . To model the density of the distribution of points in the space, use Normal distribution is very important in statistics and are most often used in the natural science to symbolize real-valued and random variables whose distributions are not known yet. The normal distribution is sometimes informally called the bell curve in the modeling stage. Thus, we need to choose big enough to fit the learning data well and small enough to avoid over fitting and other possible issues.

Network Comparisons

We compare the global network properties of our model and the PPI data networks: the degree distribution, average clustering coefficient and the diameter to find out charesteristic feature of the network of the gene and protein. However, these properties do not themselves tell us much about the structure of a network because of the difference in local structure. For example, it is insignificant to construct two networks with the same degree of distribution but with immensely different local structures. The same is also true for the clustering coefficient and the average diameter. An additional problem with using these network features for evaluating the functions of network models is the noise and incompleteness that are found in PPI networks. For example, if we take two networks with the same structure, by eleminiting edges from them we will reduce the degree of the nodes, the clustering coefficients and average diameters of networks and therefore it is found out that while some structure between the two networks might have remained similar, all three of the above properties (the degree distribution, average clustering coefficient and the diameter) might differ substantially. Many network models have earlier been introduced attempting to find out specific sets of PPI network properties, or to imitate the way in which these networks might have evolved. However, it is quite difficult to say to what extent network properties really explain network structure. Therefore, network models should use all of the available real network information, i.e., the overall network connectivity information, not only some of the network properties, to learn the structure of PPI networks. Using our network and previous observations that geometric random graphs provide a good model for PPI networks The problem of constructing a well-fitting model for PPI networks to a standard machine learning problem of density estimation have been reduced . Our model uses parts of PPI networks of high quality to learn the network structure and uses this learned knowledge to obtain model networks with random numbers of nodes and edges. Vitiligo co-occurs with other autoimmune mediated diseases, importantly Pernicious anemia, Type 1 Diabetes, and Addison's disease. The basic cause is unknown, but there is a strong genetic component, and it is multi-factorial. *P53* and its senescent-associated target genes *ICOS* and *CTLA4* were analyzed .

METHODOLOGY

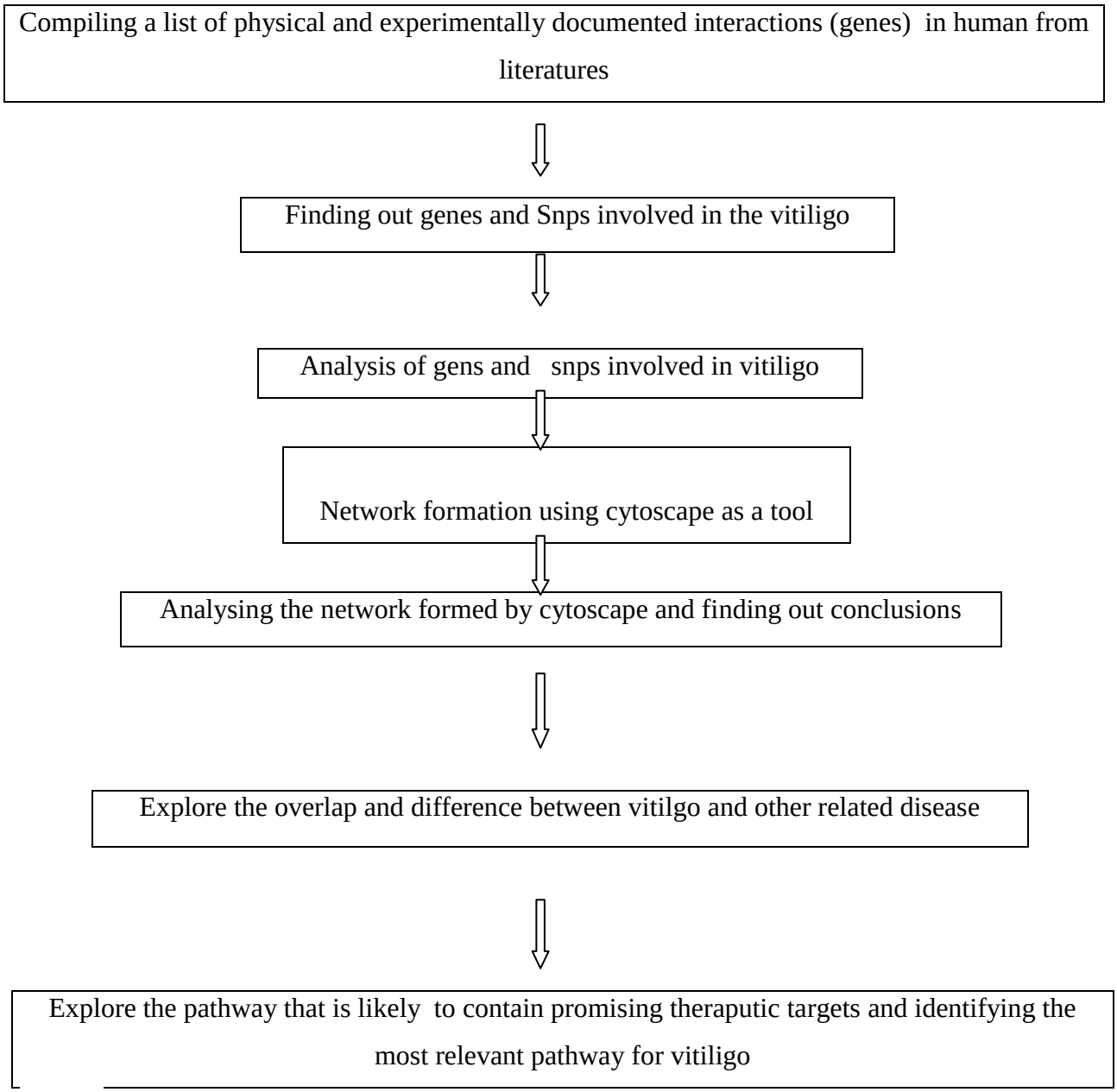
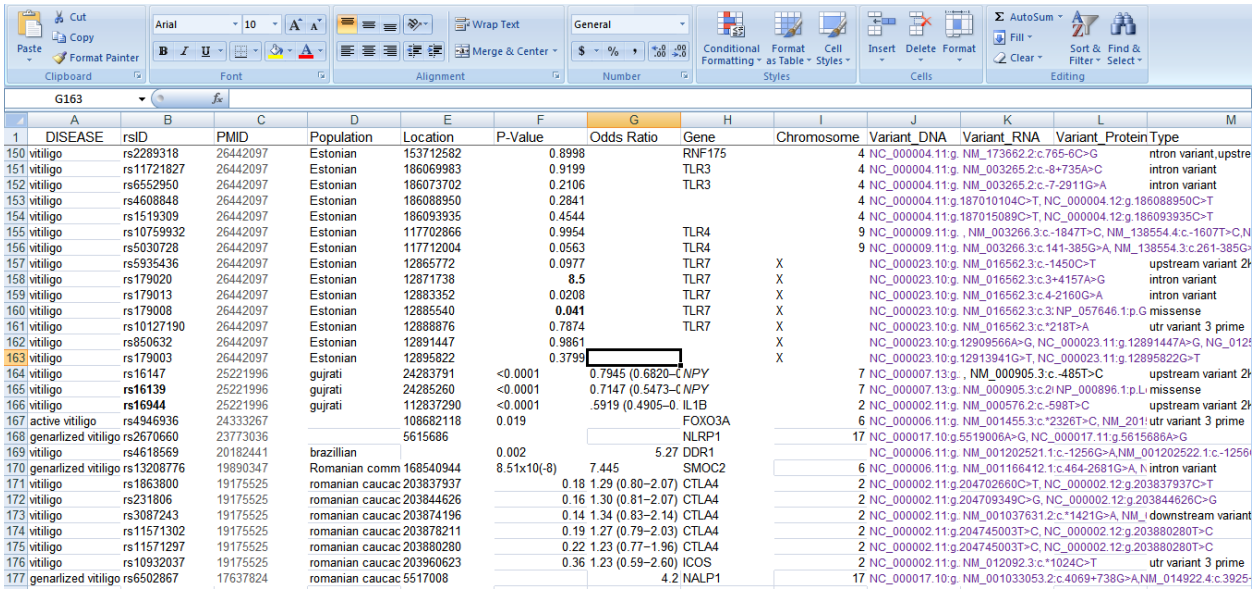


Fig 2 : Methodology used for finding out factors involved in vitiligo.

For study of interactome the first priority is to collect the data from various sources here in this process data is collected from various research papers published in pubmed and many other literatures. It would be better to find out distinguishing feature of genes involved in vitiligo so It is beter to looked upon the parameter which could be used in interactome study . The parameter used for making interactome are : Gene name , pmid , p-value, odds-ratio , rsid of snp , chromosome number , types of mutation , DNA and RNA variant , location , population . Data collection is the most important step in study of interactome as it provide information in large amount which could be used to find out some important conclusions about the interactome or pathway for a particular disease. Format of data collection is in excel sheet and its overall format and figure is given below.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DISEASE	rsID	PMID	Population	Location	P-Value	Odds Ratio	Gene	Chromosome	Variant_DNA	Variant_RNA	Variant_Protein	Type
150	vitiligo	rs2289318	26442097	Estonian	153712582	0.8998		RNF175	4	NC_000004.11:g.	NM_173662.2:c.765-6C>G		ntron variant,upstre
151	vitiligo	rs11721827	26442097	Estonian	186069983	0.9199		TLR3	4	NC_000004.11:g.	NM_003265.2:c.-8+735A>C		intron variant
152	vitiligo	rs6552950	26442097	Estonian	186073702	0.2106		TLR3	4	NC_000004.11:g.	NM_003265.2:c.-7-2911G>A		intron variant
153	vitiligo	rs4608848	26442097	Estonian	186088950	0.2841			4	NC_000004.11:g.187010104C>T,	NC_000004.12:g.186088950C>T		
154	vitiligo	rs1519309	26442097	Estonian	186093935	0.4544			4	NC_000004.11:g.187015089C>T,	NC_000004.12:g.186093935C>T		
155	vitiligo	rs10759932	26442097	Estonian	117702866	0.9954		TLR4	9	NC_000009.11:g.	, NM_003266.3:c.-1847T>C,	NM_138554.4:c.-1607T>C,N	
156	vitiligo	rs5030728	26442097	Estonian	117712004	0.0563		TLR4	9	NC_000009.11:g.	NM_003266.3:c.141-385G>A,	NM_138554.3:c.261-385G>	
157	vitiligo	rs5935436	26442097	Estonian	12865772	0.0977		TLR7	X	NC_000023.10:g.	NM_016562.3:c.-1450C>T		upstream variant 2
158	vitiligo	rs179020	26442097	Estonian	12871738	**8.5**		TLR7	X	NC_000023.10:g.	NM_016562.3:c.3+4157A>G		intron variant
159	vitiligo	rs179013	26442097	Estonian	12883352	0.0208		TLR7	X	NC_000023.10:g.	NM_016562.3:c.4-2160G>A		intron variant
160	vitiligo	rs179008	26442097	Estonian	12885540	**0.041**		TLR7	X	NC_000023.10:g.	NM_016562.3:c.3:	NP_057646.1:p.G	missense
161	vitiligo	rs10127190	26442097	Estonian	12888876	0.7874		TLR7	X	NC_000023.10:g.	NM_016562.3:c.*218T>A		utr variant 3 prime
162	vitiligo	rs850632	26442097	Estonian	12891447	0.9861			X	NC_000023.10:g.12909566A>G,	NC_000023.11:g.12891447A>G,	NG_012	
163	vitiligo	rs179003	26442097	Estonian	12895822	0.3799			X	NC_000023.10:g.12913941G>T,	NC_000023.11:g.12895822G>T		
164	vitiligo	rs16147	25221996	gujrati	24283791	<0.0001	0.7945 (0.6820–(	*NPY*	7	NC_000007.13:g.	, NM_000905.3:c.-485T>C		upstream variant 2
165	vitiligo	**rs16139**	25221996	gujrati	24285260	<0.0001	0.7147 (0.5473–(	*NPY*	7	NC_000007.13:g.	NM_000905.3:c.2	NP_000896.1:p.L	missense
166	vitiligo	**rs16944**	25221996	gujrati	112837290	<0.0001	.5919 (0.4905–0.	IL1B	2	NC_000002.11:g.	NM_000576.2:c.-598T>C		upstream variant 2
167	active vitiligo	rs4946936	24333267		108682118	0.019		FOXO3A	6	NC_000006.11:g.	NM_001455.3:c.*2326T>C,	NM_201	utr variant 3 prime
168	genarlized vitiligo	rs2670660	23773036		5615686			NLRP1	17	NC_000017.10:g.5519006A>G,	NC_000017.11:g.5615686A>G		
169	vitiligo	rs4618569	20182441	brazillian		0.002	5.27	DDR1		NC_000006.11:g.	NM_001202521.1:c.-1256G>A,	NM_001202522.1:c.-1256	
170	genarlized vitiligo	rs13208776	19890347	Romanian comm	168540944	8.51x10(-8)	7.445	SMOC2	6	NC_000006.11:g.	NM_001166412.1:c.464-2681G>A,	N	intron variant
171	vitiligo	rs1863800	19175525	romanian caucac	203837937	0.18	1.29 (0.80–2.07)	CTLA4	2	NC_000002.11:g.204702660C>T,	NC_000002.12:g.203837937C>T		
172	vitiligo	rs231806	19175525	romanian caucac	203844626	0.16	1.30 (0.81–2.07)	CTLA4	2	NC_000002.11:g.204709349C>G,	NC_000002.12:g.203844626C>G		
173	vitiligo	rs3087243	19175525	romanian caucac	203874196	0.14	1.34 (0.83–2.14)	CTLA4	2	NC_000002.11:g.	NM_001037631.2:c.*1421G>A,	NM_	downstream variant
174	vitiligo	rs11571302	19175525	romanian caucac	203878211	0.19	1.27 (0.79–2.03)	CTLA4	2	NC_000002.11:g.204745003T>C,	NC_000002.12:g.203880280T>C		
175	vitiligo	rs11571297	19175525	romanian caucac	203880280	0.22	1.23 (0.77–1.96)	CTLA4	2	NC_000002.11:g.204745003T>C,	NC_000002.12:g.203880280T>C		
176	vitiligo	rs10932037	19175525	romanian caucac	203960623	0.36	1.23 (0.59–2.60)	ICOS	2	NC_000002.11:g.	NM_012092.3:c.*1024C>T		utr variant 3 prime
177	genarlized vitiligo	rs6502867	17637824	romanian caucac	5517008		4.2	NALP1	17	NC_000017.10:g.	NM_001033053.2:c.4069+738G>A,	NM_014922.4:c.3925-	

Fig3: Data obtained after collecting from various sources.

For the data assortment, approximately 150 research papers were searched from various journal mostly from pubmed and went through several processing work to get the final data collection. Using this data we can find lots of conclusion related to network , connectivity , relationship between genes and protein interaction. This data could be useful in gathering gene-gene interaction , protein- protein interaction , protein- gene interaction. The data contain snp`s

involved in gene malfunctioning which leads to the formation of faulty protein. So using this data source we can get wide variety of results and conclusion which could enhance our understanding on vitiligo and its causing factors .

RESULTS AND ANALYSIS

For study of interactome for determinig novel pathways and genes of vitiligo a set of 90 genes and their respective parameters and features are taken to get result for the analysis of genes related to vitiligo some of the basic features which are used to get the desired results are **name of the gene , snps related to the genes , PMID , p-value and odds ratio**. Various tools have been used to find out respective relationship between genes and their respective protein. These tools are:

1) cytoscape (http://www.cytoscape.org/)

2) string (http://string-db.org/)

3) Gorilla (http://cbl-gorilla.cs.technion.ac.il/)

4) David bioinformatics resources (https://david.ncifcrf.gov/)

5) dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)

6) NCBI (http://www.ncbi.nlm.nih.gov/)

7) KEGG pathway database (http://www.genome.jp/kegg/pathway.html)

8) PANTHER classification system(http://www.pantherdb.org/)

Gorilla gene ontology

GOrilla is a bioinformatics tool for identifying and visualizing enriched GO(gene ontology) terms in ranked lists of genes. user should supply a single list of genes where they are ranked according to some biological measurement (for example expression level or using some other parameters). The software will search for GO terms that are enriched in the top of the list of the

genes compared to the rest of the list using the mHG statistics. Note that the list should not be unnecessary, i.e. each gene should appear in the list only once at a time .gorilla gene ontology contain database and properties of genes of many species. On inserting the list of genes in column 1, following pathway which shows various biological processes for which genes are responsible was obtained . Using this It is concluded that list of genes is mainly responsible for some biological processes such as response to oxygen level , response to stress , response to hypoxia .Here in the given figure below various categories of genes and their respective biological process is shown.

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0001666	response to hypoxia	4.25E-4	1E0	5.49 (79,6,12,5)	[-] Hide genes MTHFR - methylenetetrahydrofolate reductase (nad(p)h) SOD3 - superoxide dismutase 3, extracellular LTA - lymphotoxin alpha CAT - catalase IL1B - interleukin 1, beta
GO:0036293	response to decreased oxygen levels	4.25E-4	5.78E-1	5.49 (79,6,12,5)	[-] Hide genes MTHFR - methylenetetrahydrofolate reductase (nad(p)h) SOD3 - superoxide dismutase 3, extracellular LTA - lymphotoxin alpha CAT - catalase IL1B - interleukin 1, beta
GO:0070482	response to oxygen levels	4.25E-4	3.86E-1	5.49 (79,6,12,5)	[-] Hide genes MTHFR - methylenetetrahydrofolate reductase (nad(p)h) SOD3 - superoxide dismutase 3, extracellular LTA - lymphotoxin alpha CAT - catalase IL1B - interleukin 1, beta

Fig 4: Categories of genes and biological process

PANTHER :

The PANTHER (Protein ANalysis THrough Evolutionary Relationship) a sorting System was designed and manufactured to classify proteins (and genes) in order to facilitate high-throughput study. Proteins have been classified according to Family and subfamily ,Molecular function .using panther tool we can find out protein formed by the genes its family and sub family. we can also get protein class of the genes. Using this database tool we can find out how many genes in the set of genes belong to what category of genes and produce which type of protein. On inserting the set of genes in the panther database result is obtained in the form of document describing categories of the genes and protein it produce.

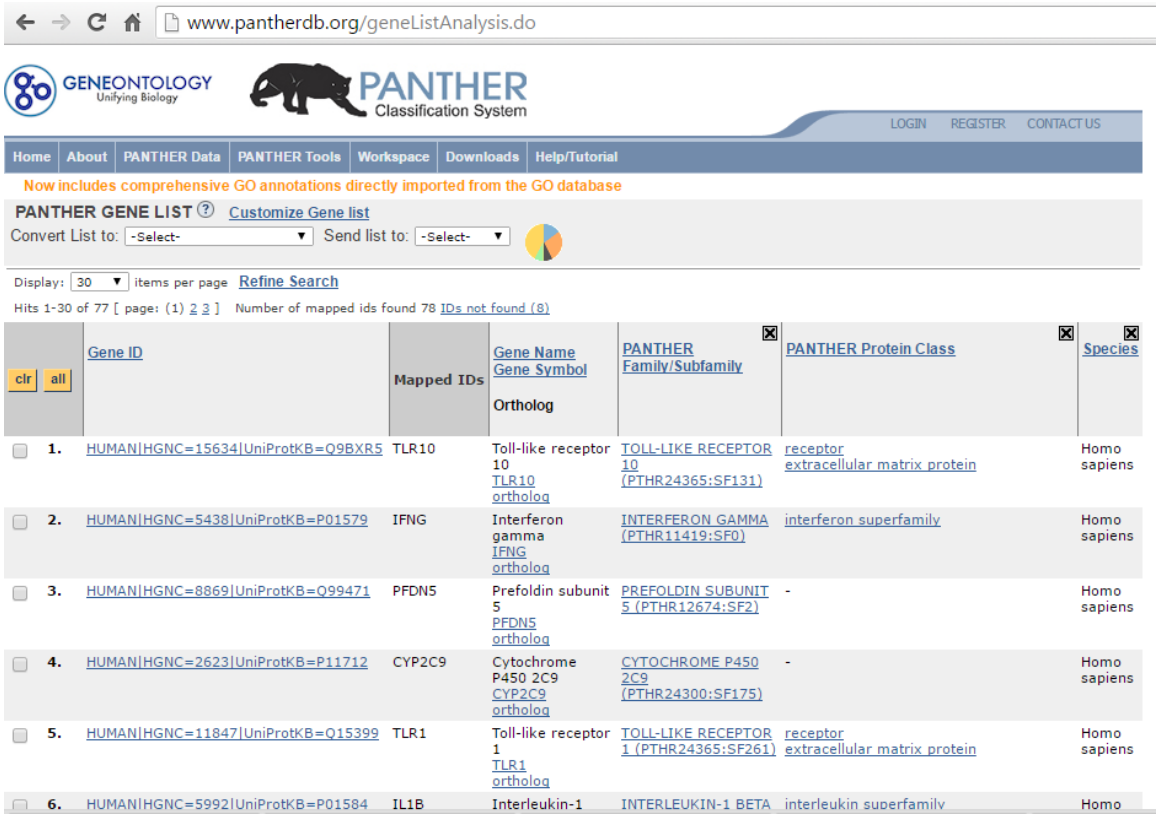


Fig5: Sheet describing the genes and its charasteristic feature.

Apart from this we can also get the result in the form of pie chart which is given below . first figure shown below shows protein class related to gene while second figure shows graph form of biological process of given gene list.

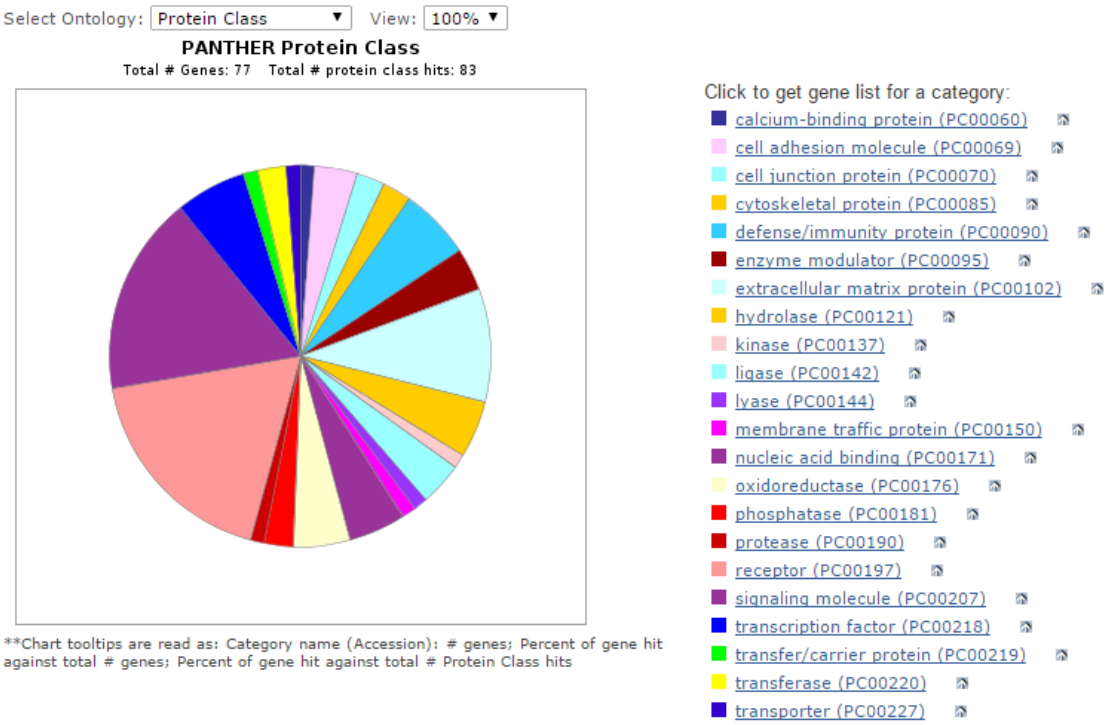


Fig 6: Result in the form of pie chart

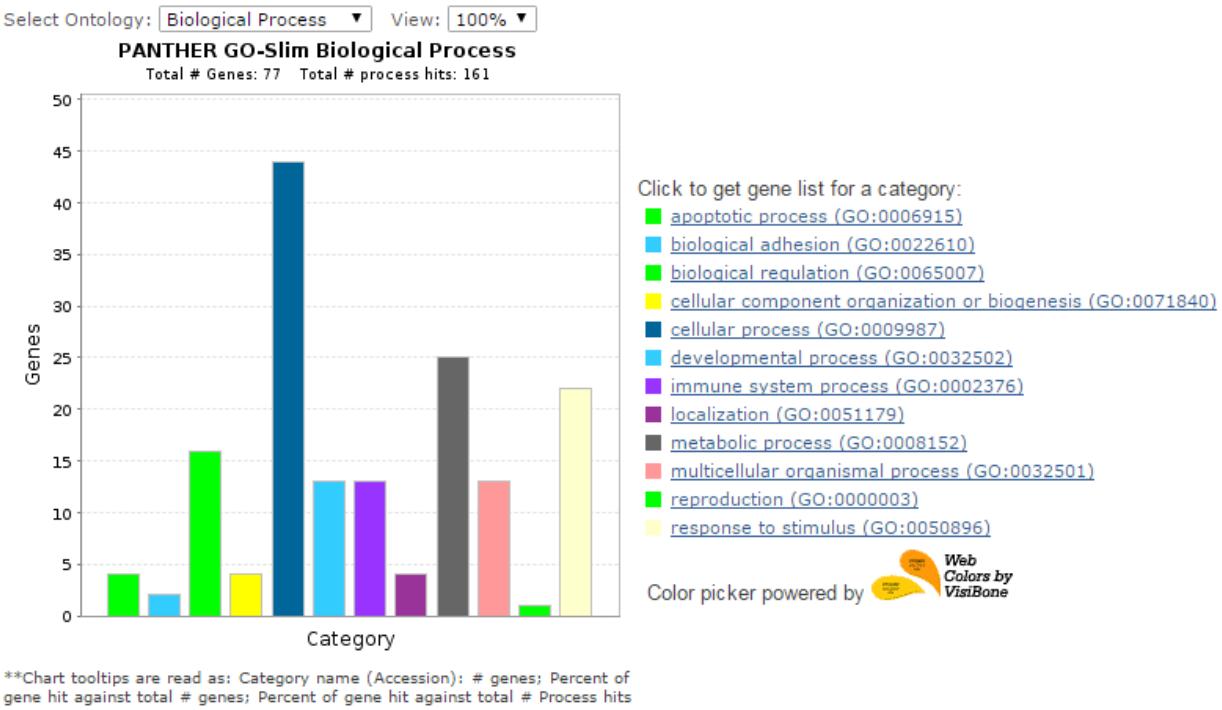


Fig 7: Biological processes of the genes in the form of bar graph.

String database

String is a database tool which is used to find out interactome of protein. On inserting the list of protein obtained from panther database, result is obtained in the form of interactome shown below.

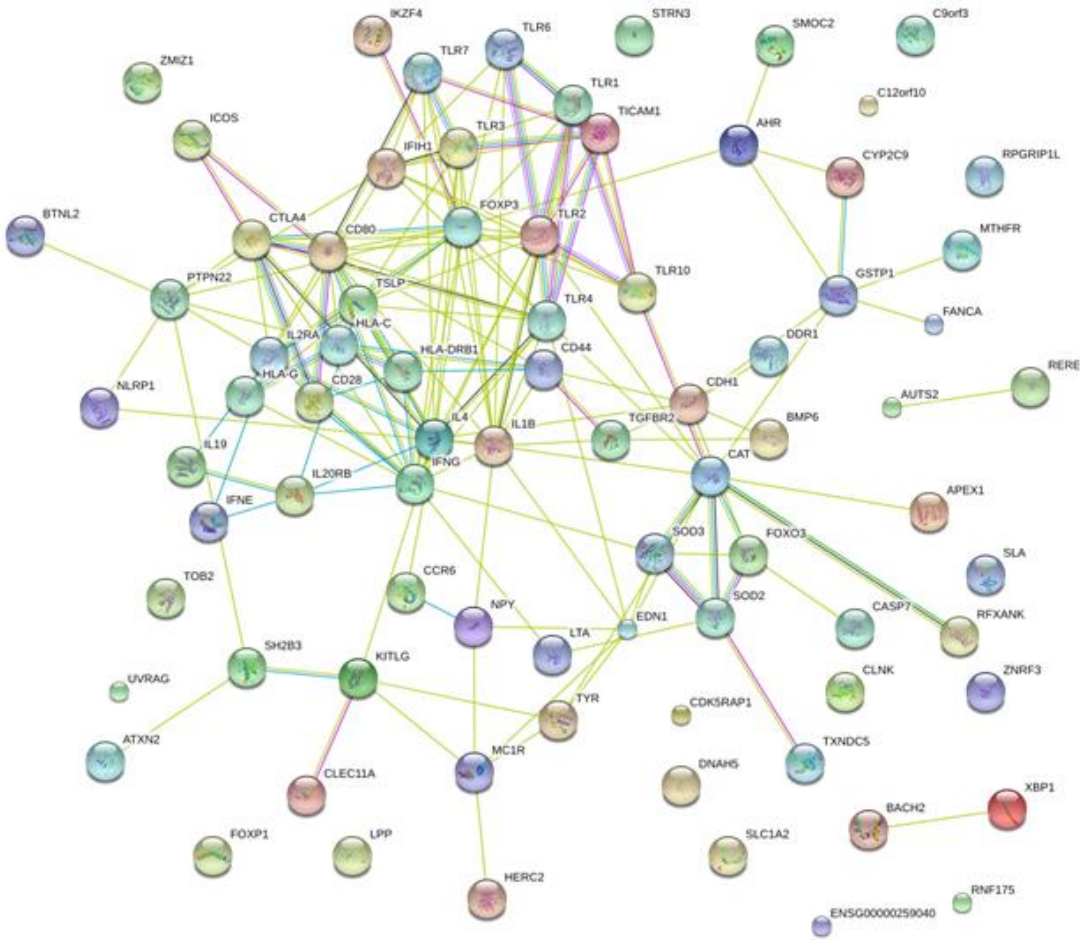


Fig 8:- Interactome by string.

In string It has been found out that 2-3 subnetwork in the interactome with maximum no. of interaction with ILRA2 , TLR2 , CTLA4, TICAM1 various color of edges indicate various process under going in the interactome. during the analysis of the interactom it is been out found that the genes involved in the interactome is also present . In general, there are many pathways in which our listed genes are present which is shown below.

Pathway ID	Pathway description	Count in network	False discovery rate
04940	Type 1 diabetes mellitus	8	7.62e-10
05330	Allograft rejection	7	7.58e-09
05332	Graft-versus-host disease	7	7.8e-09
04620	Toll-like receptor signaling pathway	9	1.51e-08
04060	Cytokine-cytokine receptor interaction	12	2.22e-08
05320	Autoimmune thyroid disease	7	3.85e-08
05168	Herepes simplex infection	10	5.4e-08
05323	Rheumatoid arthritis	8	5.4e-08
05321	Inflammatory bowel disease	7	1.72e
04514	Cell adhesion molecules	8	1.57e-06
05164	Influenza A	8	5.97e-06
05152	Tuberculosis	8	6.43e-06
04672	Intestinal immune network for IgA production	5	1.39e-05
04640	Hematopoietic cell lineage	6	1.71e-05
04630	Jak-stat signaling pathway	7	3.63e-05
05416	Viral myocarditis	5	3.63e-05
04612	Antigen processing and presentation	5	8.57e-05
04660	Tcell receptor signaling pathway	5	0.000514
05166	HTLV-1 infection	7	0.000639
04668	TNF signaling pathway	4	0.00714
05169	Epstein-Barr virus infection	5	0.00762

Table 1 : Genes in the interactome responsible for number of diseases

Cytoscape

Cytoscape (http://www.cytoscape.org/) is an open source bioinformatics software platform or tool for visualize molecular interaction networks and integrating with gene expression profile and other state data . Cytoscape is a widely used tool for studying interaction and relationship between various genes and protein. There are 150 plugins which can be used to find out various types of interaction to have wide variety of knowledge about interactions using many parameters. Plugins may be developed using the Cytoscape open Java software design by anyone and plugin community development is expectant

Normal network formation

Figure shown below denotes normal network formed on taking gene list 1 as source list 2 as a target and list 3 as a interaction value using all these value i got the network formed as given below in organic layout. It also shows two self loop. Here in this case i take p-value as a interaction value. Here in this case network is just showing relationship between nodes and their mutual connectivity which is mostly biennial in nature

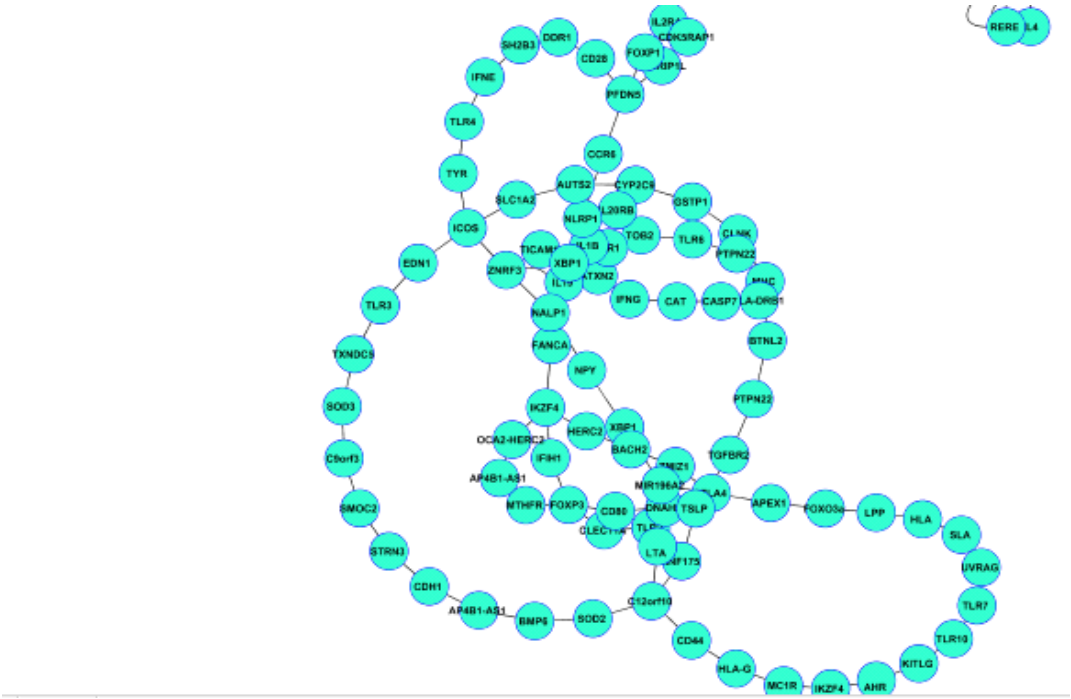


Fig 9 : normal network formation using cytoscape.

Different type of network is formed when p value or odds ratio is not used as a interaction value. It is shown below. Three or four big loops in the interactome shows that the genes involved in this disease must be a part form some large pathways. This was confirmed in the result of gorilla gene ontology which shows that the genes obtained is mainly related to 3- 4 biological processes such as response to decreased oxygen level , response to oxygen level , response to hypoxia.

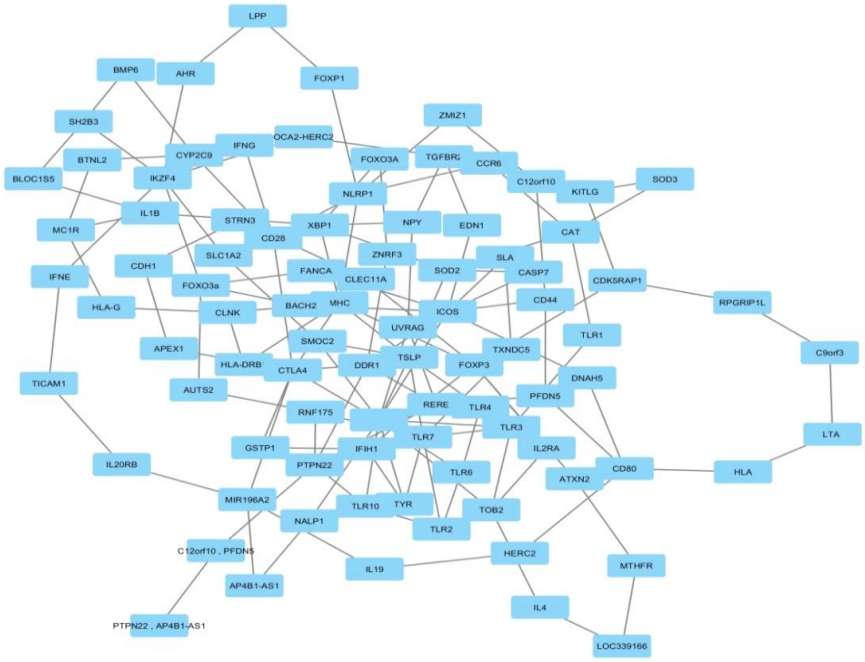


Fig10 : Figure shows connectivity between the nodes which are connected two or more nodes.

Here it has been found out that some genes with maximum degree more than 7 or 8 could act as a hub protein. These genes are FOXP3 , CTLA4 , ICOS , MHC, TYR. The figure given below shows genes with maximum no. Of degrees more than7 or 8.

Name	Degree	Betweenness centrality	Closeness centrality	Stress cetrality
FOXP3	6	0.14017928	0.28434505	2004
MHC	7	0.22320587	0.33458647	3918
CTLA4	7	0.24367388	0.34765625	4134
ICOS	9	0.18852629	0.32246377	3170

Table 2: Table above shows the characteristic feature of four main genes with its degree , betweenness centrality , closeness centrality , stress centrality.

Simple charesteristic result of the network

In the figure given below it has been shown various feature of the network formed such as clustering coefficient , connected component , network diameter , network radius , shortest path , charesteristic pathlength , average number Of neighbours , number Of nodes , network density , isolated nodes , number Of self loops , multi edge node pair , analysis time(sec).

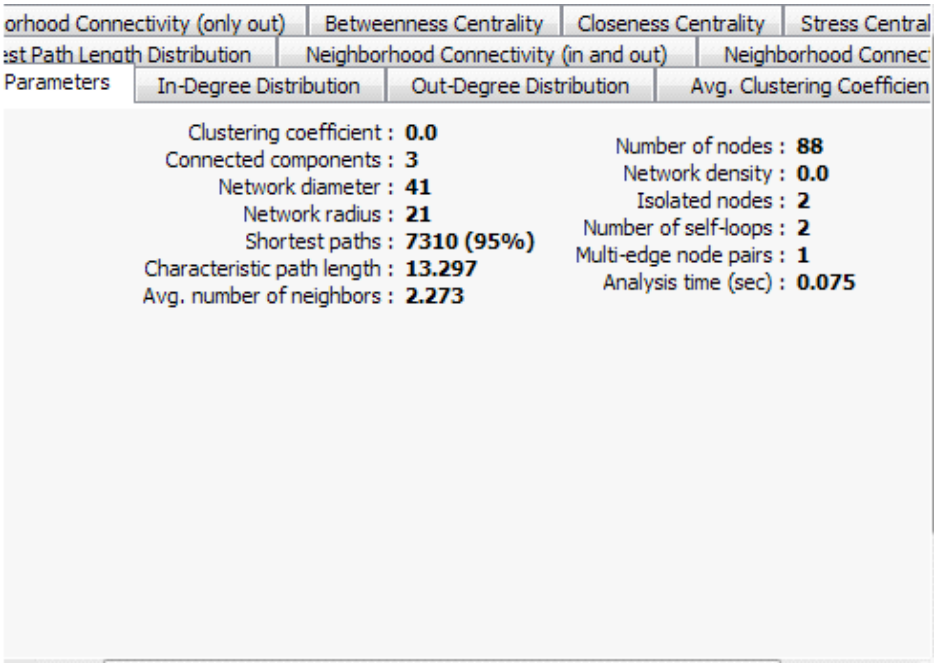


Fig 11: Characteristic features of the network.

The above figure provides a very primary information of the network in which we can find out how the network look in the form of coonection showing nodes and edges. Using this primary information we can judge number of connections , number of nodes, how each node is connected to the other , what is the degree of each node and many more informations..using this primary information we can find out number of clusters present in the network . In this case clustering coefficient of 0 shows that there is no cluster formation took place. So this primary information could be very handy data in concluding our result.

Network analyzer - plot parameter

Here in this part comparative analysis between eccentricity and out degree of the network the eccentricity, denoted by “e” parameter associated with every conic section has been found out. It can be thought of as a measure of how much the conic section deviates from being circular .

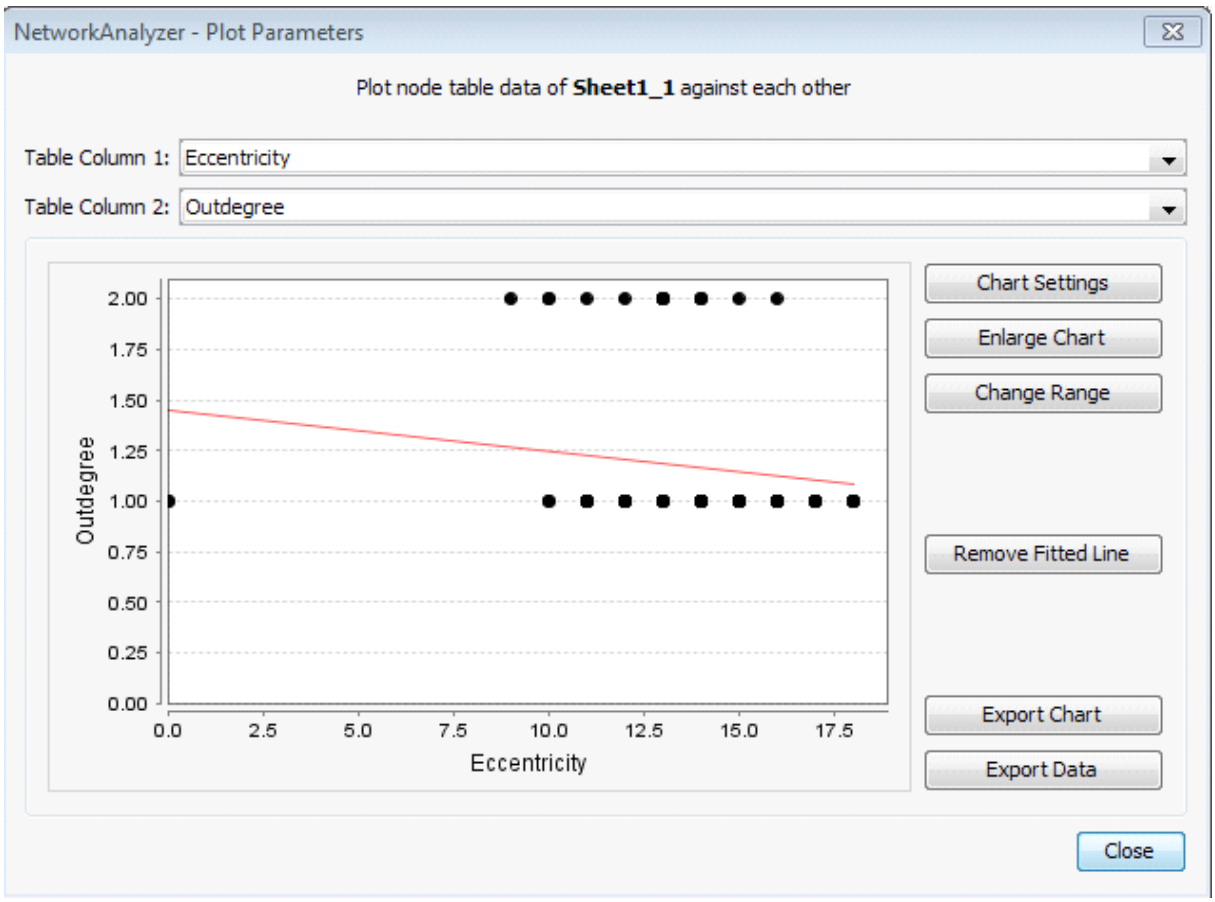


Fig 12: Network analyzer – plot parameter graph

Betweeness centrality

Betweeness centrality is a indicator of a node's centrality in a network structure. It is a association between number of neighbour and betweeness centrality of genes and related genes .It is equal to the number of shortest paths from all vertices to all others that pass through that node. It is an indicator how genes are related to central gene or source gene. In the graph given below maximum value -1 minimum value- 0 has been shown . In the graph given below it has been shown that betweeness centrality ranges from .04 to 0.35 which shows centrality is approximately 35%.

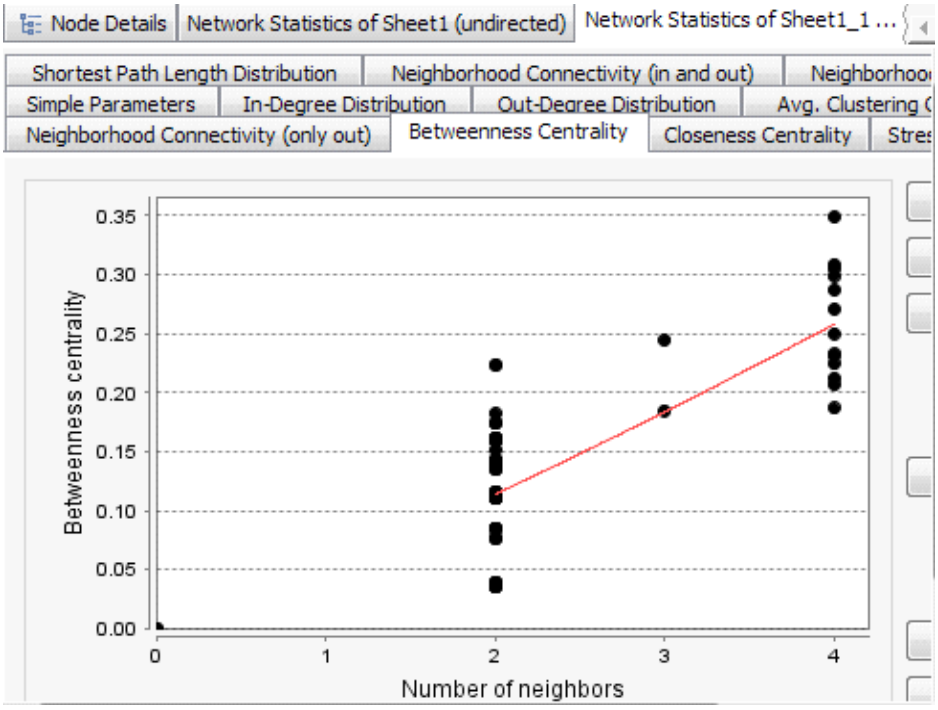


Fig 13: Graph showing betweenness centrality and no. of neighbor.

Shortest path length distribution

Shortest path length distribution facilitates the quick transfer of information . In network also genes on producing protein desires to transfer its information through the shortest path . Here also during the procedure it is found out that shortest path length of the network using comparative analysis between frequency and path length is 15 at maximum frequency of 450.

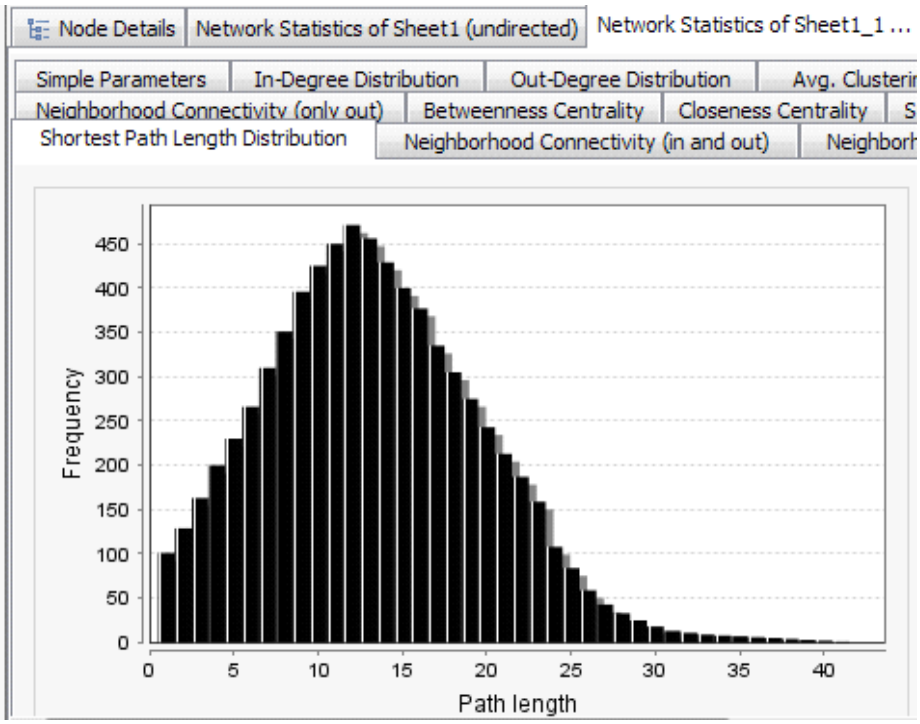


Fig 14: Graph showing relationship between frequency and path length for the calculation of shortest path length distribution.

Stress centrality

The stress centrality of a node “n” is the number of shortest paths passing through n. A node has a high stress if it is passes through high number of shortest paths. This parameter is defined only for networks without multiple edges or multiple nodes. The stress centraility distribution gives the number of nodes with stress s for different values of s. The values for the stress are grouped into bins whose size grows exponentially by a factor of 10 Here in this figure it is indicated that most of the nodes have stress centrality of 100.

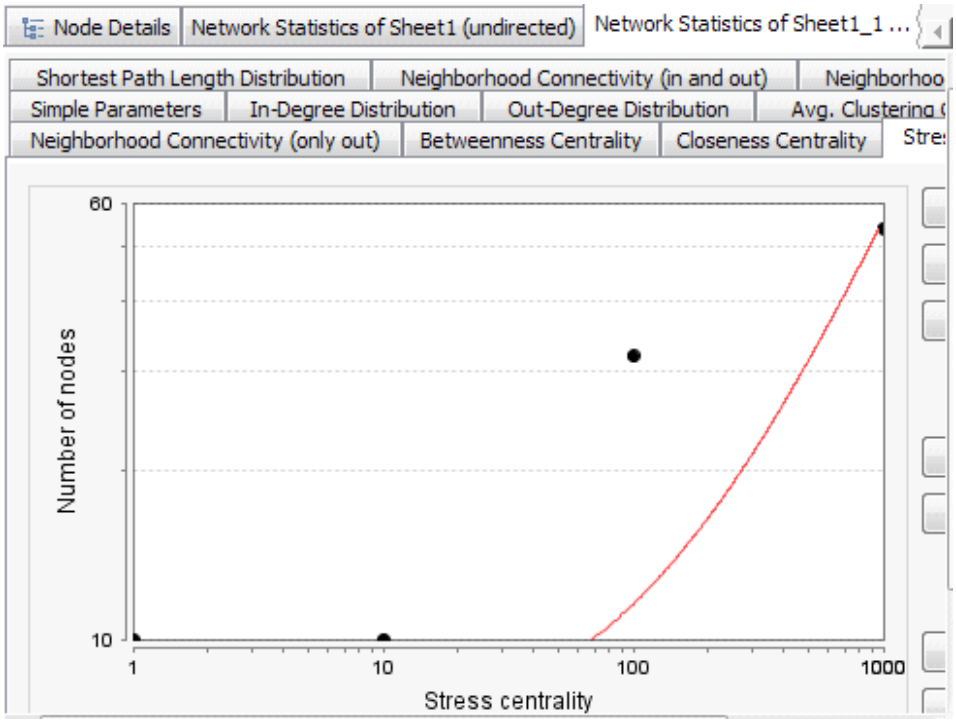


Fig 15: graph showing relationship between number of nodes and stress centrality.

Strees centrality of a node allow us to find the amount of control that this node apply over the interactions of other nodes in the same network This measure is beneficial for the nodes that join communities (dense subnetworks), rather than the nodes that lie inside a community. Network Anlayzer uses fast algorithm by Brandes for the computation of node betweenness centrality.

This algorithm shows its complexity in the form of O(NM), N being the number of nodes and M the number of edges in the network. During the analysis it has been found that density is quite low because there are small number of genes that has been obtained.

Neighbourhood connectivity (in and out)

Here in the figure given below relationship between average connectivity and number of neighbours has been shown . It shows that average connectivity of 3 is maximum.

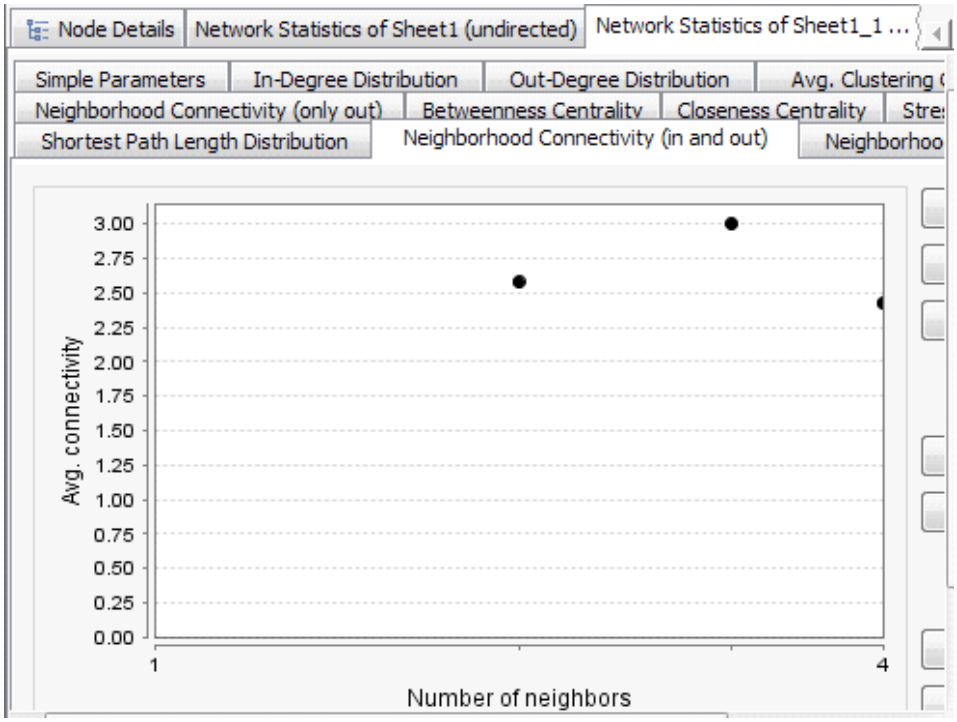


Fig 16: Graph showing relationship between avg. connectivity and number of neighbour for the calculation of neighbourhood connectivity.

The connectivity between the nodes can be defined as the number of its neighbours. The neighbourhood connectivity of a node "n" is defined as the average connectivity of all neighbours of "n" i.e how a particular node is connected to other node . The neighbuorhood connectivity distribution gives the average of the neighborhood connectivities of all nodes "n"

with “k” neighbors. Here in the project work it is found out that the average neighbourhood connectivity is 2.5 , 3.0. It means that on an average each node is connected to three other nodes.

M-code result

In the network analysis part of the genes it has been found that there is no cluster exists which shows that the genes involved in the vitiligo is a part of some other pathways which could be the possible reason for not forming any cluster.

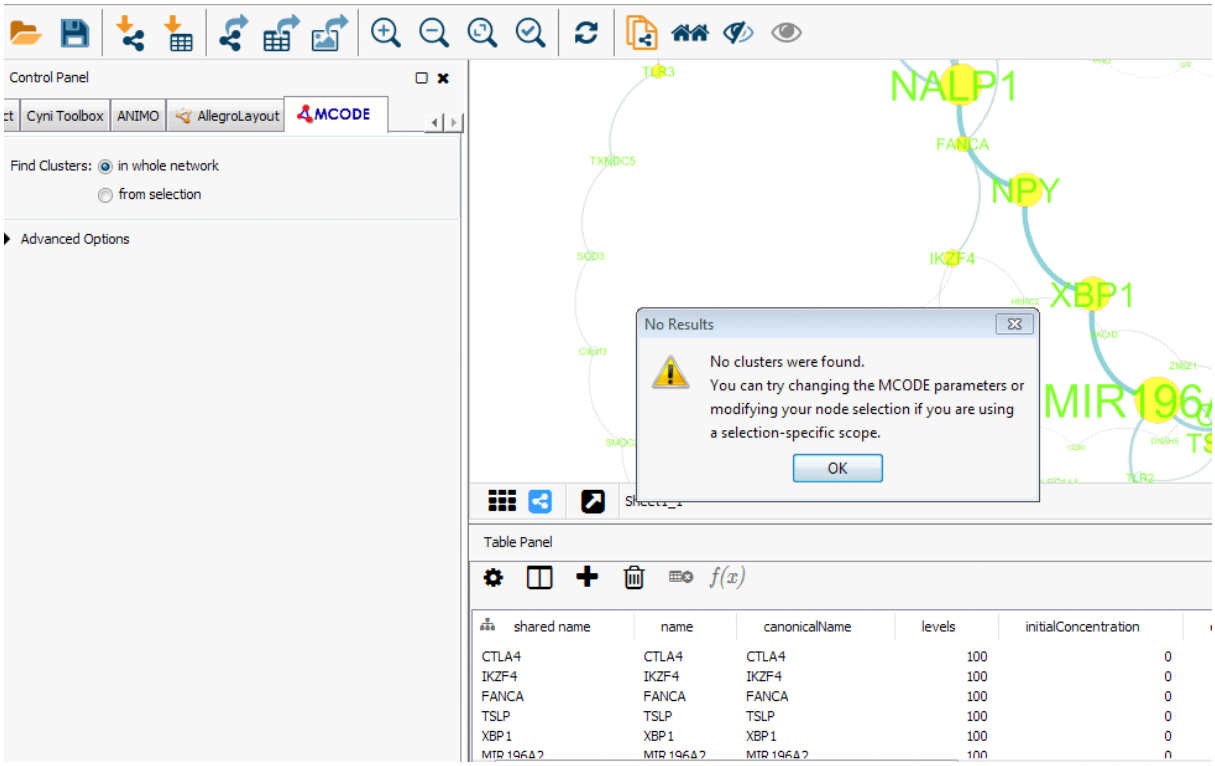


Fig 17: Result of M-code connectivity.

Pathway scoring.

Using cytoscape a plugin called as pathway scoring is installed and the known genes were tested Using this pathway we can find out genes which are involved in KEGG pathway. On loading the gene list with the parameter used in formation of network ,kegg pathway of genes as shown below is obtained. Here we can see that 4 to 5 clusters get formed on finding clusters based on pathway scoring

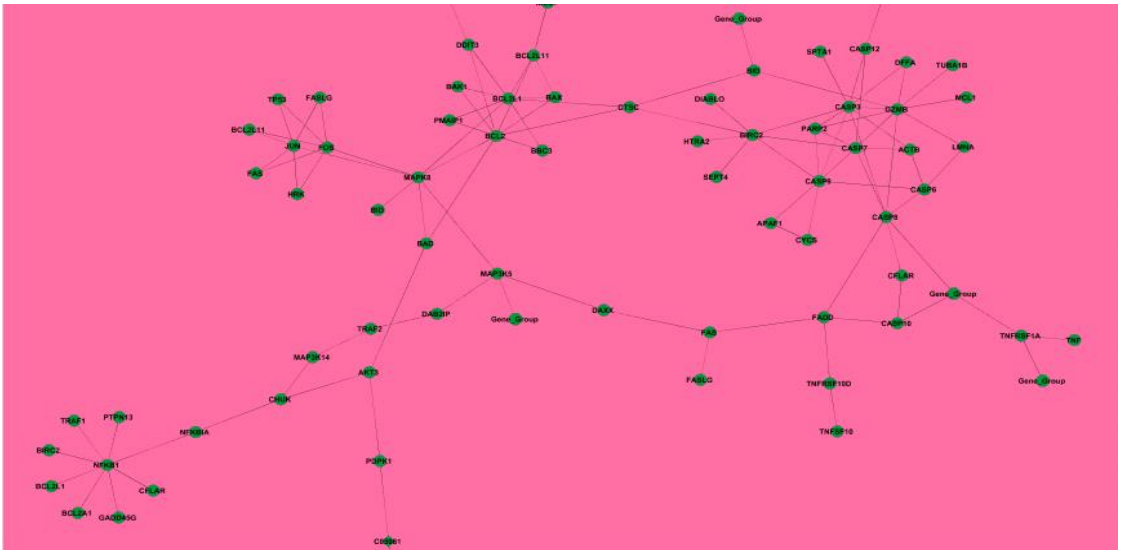


Fig 18: Result showing pathway scoring.

Here it is also found out that there are some genes which are not connected to any other genes. Three radially connected genes were also found out. some of the pathways in which given genes are involved theses patways are shown below.

1) cytokine-cytokine receptor interaction

2) NF- kappa signalling pathway

3) MAPK signalling pathway

4) protein processing in endoplasmic reticulum

5) p53 signalling pathway.

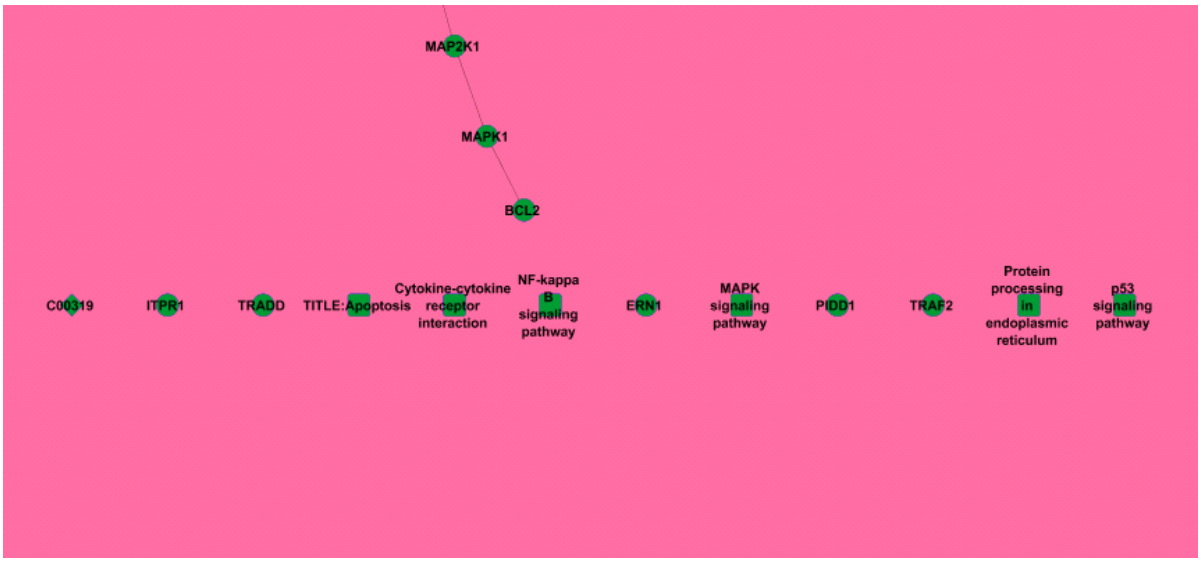


Fig 19: Figure showing pathway scoring.

CLUE GO result.

On adding genes result has been found out in the form of molecular description and biological description as given below in which each genes along with its respective biological description has been shown.

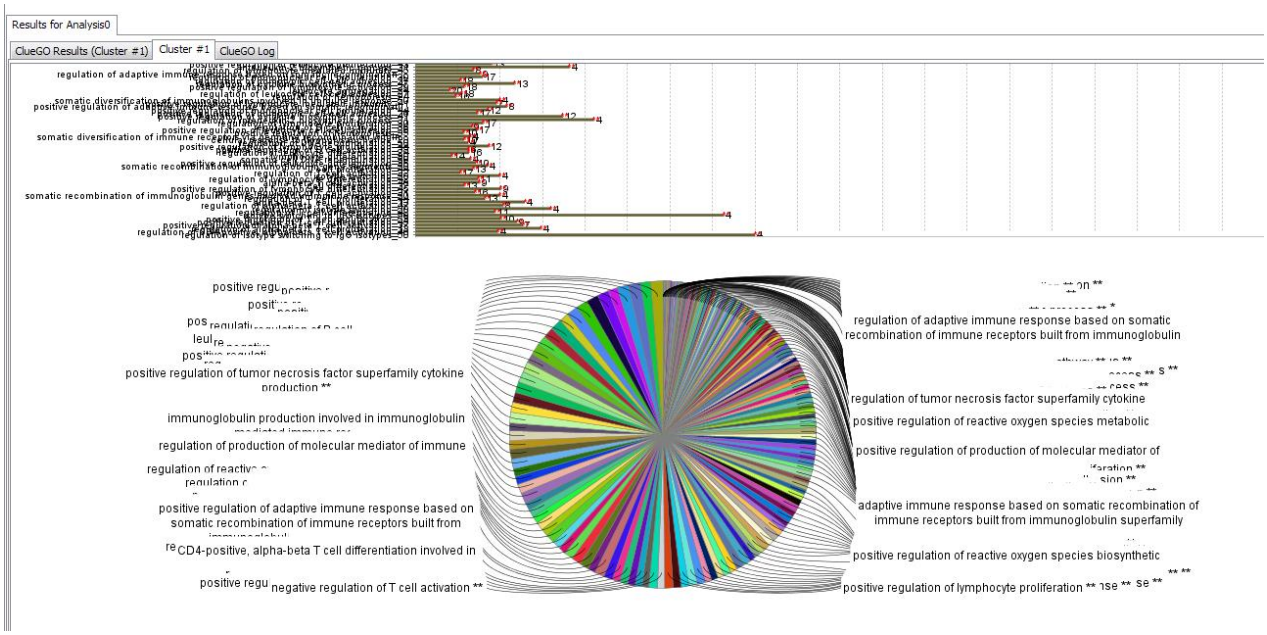


Fig 20 : Result of clugo showing categories of genes.

Result of clugo:

Genes in KEGG	6985
Genes in interPro-proteinDomains	12709
Genes in GO ImmuneSystemProcess	2815
Genes in REACTOME	8402
Genes in GO Molecular function	17045
Genes in GO cellular component	18114
Genes in GO-Biological process	16950
Total number of Genes in clusters	4.71%
All the genes found in initial cluster 1	96.3%
Genes found from all clusters after selection	57(70.37%)
Kappa score grouping	221 groups
Final Kappa score groups	221
Terms not grouped	0
Number of GO all term specific for cluster	1:418
GO-Biological process-GOA	16h18
GO-Cellular component-GOA	16h18
GO-Immune System process-GOA	16h18
GO-Molecular Function-GOA	16h18
Correction method used	Bonferroni step down
Min GO level	3
Max GO level	8
Min percentage	4%

Over view term	Smallest P value
Sharing group %	50 %
Sample file name	Network selection
Evidence code used	All

Table 3: Result of clugo gene analysis

kappa score :

kappa score coefficient is a statistic which measures agreement for qualitative (categorical) items. It is generally thought to be a more vigorous measure than simple percent agreement calculation, since κ takes into account the agreement occurring by chance. It gives a score of how much homogeneity, or consensus. It is calculated by the formula .

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

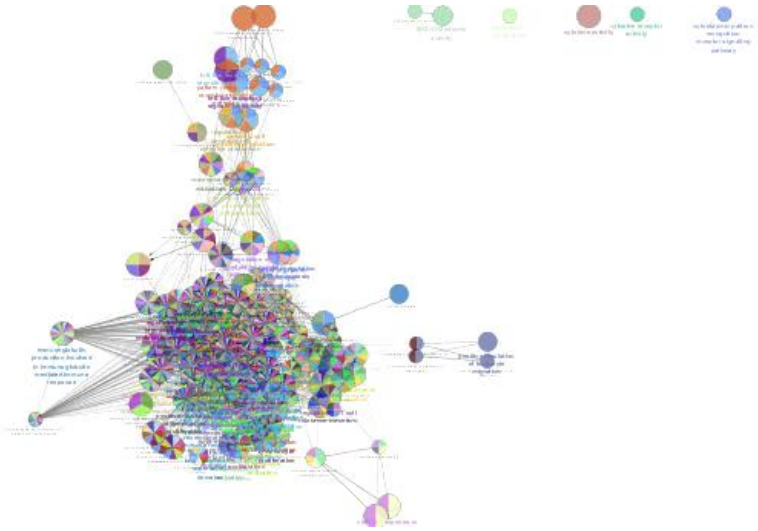


Fig 21: Figure shows clustering of the genes in the clugo analysis

Hierarchial clustering.

Here in this diagram hierarchial clustering of the genes have been shown where it has been found out that **DNAH5** is present at the lowest level of hierarchy.

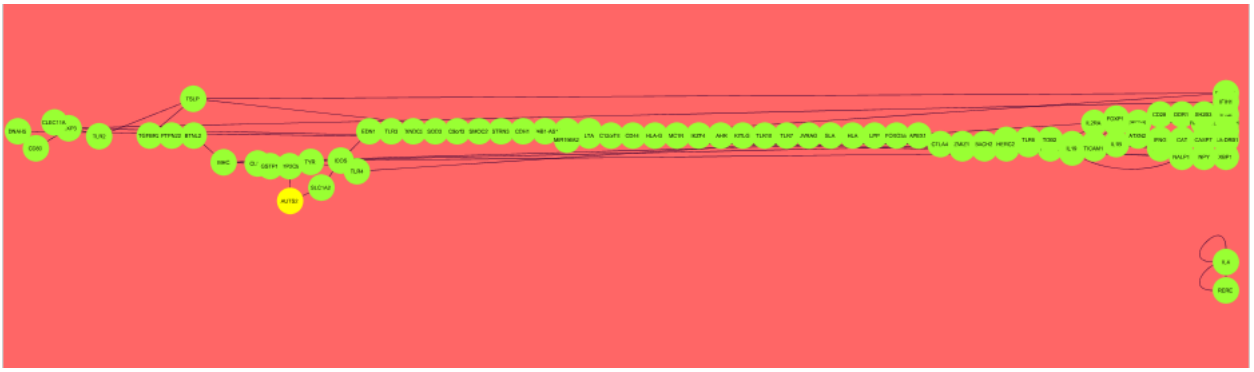


Fig 22: Figure showing hierarchial clustering of genes.

DAVID Functional Annotation Tool

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an update to the sixth version of original web-accessible programs. DAVID provides a inclusive set of functional annotation tools for investigators to find out biological meaning behind large list of genes. For any given gene list, DAVID tools can be used in : Identify large enriched biological themes, particularly GO terms , Discover enriched functional-related gene groups, Cluster redundant annotation terms, envisage genes on BioCarta & KEGG pathway maps, Display related many-genes-to-many-terms on 2-D view., finding out other functionally related genes which are not present in the list and List interacting proteins etc.

David result.

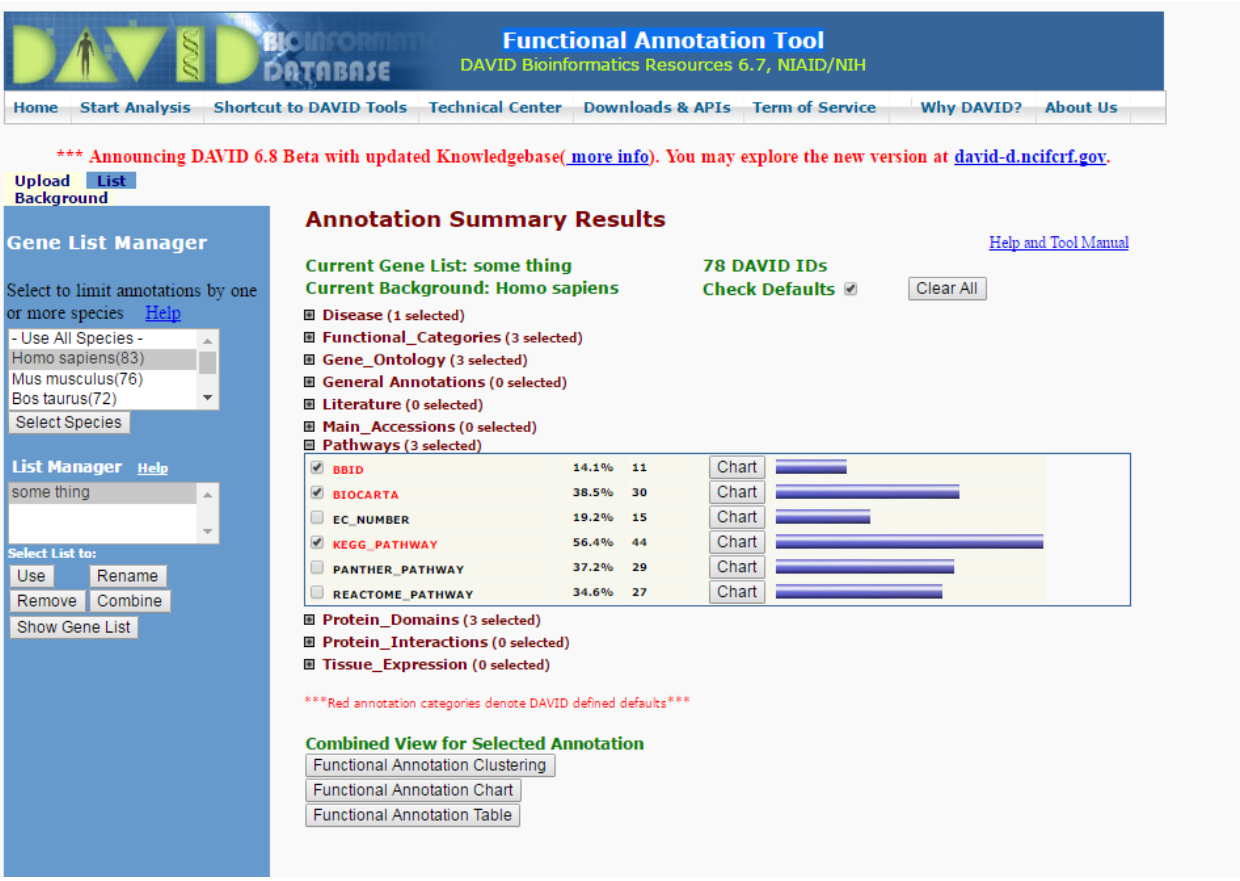


Fig 23: Annotation summary of the DAVID database tool of homo sapiens.

Diagram below shows the pathways in which genes involved . there were three pathways in which our genes were found out these were BBID , BIOCARTA , KEGG pathway.

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

*** Announcing DAVID 6.8 Beta with updated Knowledgebase(more info). You may explore the new version at david-d.ncifcrf.gov.

Functional Annotation Table

Help and Manual

Current Gene List: some thing
Current Background: Homo sapiens
938 DAVID IDs

44 record(s)

Download File

Mthfr	5,10-methylenetetrahydrofolate reductase (NADPH)	Related Genes	Homo sapiens
KEGG_PATHWAY	One carbon pool by folate, Methane metabolism,		
Apex1	APEX nuclease (multifunctional DNA repair enzyme) 1	Related Genes	Homo sapiens
KEGG_PATHWAY	Base excision repair,		
CD28	CD28 molecule	Related Genes	Homo sapiens
KEGG_PATHWAY	Cell adhesion molecules (CAMs), T cell receptor signaling pathway, Intestinal immune network for IgA production, Type I diabetes mellitus, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease, Viral myocarditis,		
CD44	CD44 molecule (Indian blood group)	Related Genes	Homo sapiens
KEGG_PATHWAY	ECM-receptor interaction, Hematopoietic cell lineage,		
CD80	CD80 molecule	Related Genes	Homo sapiens
KEGG_PATHWAY	Cell adhesion molecules (CAMs), Toll-like receptor signaling pathway, Intestinal immune network for IgA production, Type I diabetes mellitus, Autoimmune thyroid disease, Systemic lupus erythematosus, Allograft rejection, Graft-versus-host disease, Viral myocarditis,		
KITLG	KIT ligand	Related Genes	Homo sapiens
KEGG_PATHWAY	Cytokine-cytokine receptor interaction, Hematopoietic cell lineage, Melanogenesis, Pathways in cancer,		
Nlrp1	NLR family, pyrin domain containing 1	Related Genes	Homo sapiens
KEGG_PATHWAY	NOD-like receptor signaling pathway,		
SH2B3	SH2B adaptor protein 3	Related Genes	Homo sapiens
KEGG_PATHWAY	Neurotrophin signaling pathway,		
BMP6	bone morphogenetic protein 6	Related Genes	Homo sapiens
KEGG_PATHWAY	Hedgehog signaling pathway, TGF-beta signaling pathway,		
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	Related Genes	Homo sapiens
KEGG_PATHWAY	Cell adhesion molecules (CAMs), Adherens junction, Pathogenic Escherichia coli infection, Pathways in cancer, Endometrial cancer, Thyroid cancer, Melanoma, Bladder cancer,		
casp7	caspase 7, apoptosis-related cysteine peptidase	Related Genes	Homo sapiens
KEGG_PATHWAY	Apoptosis, Alzheimer's disease,		
cat	catalase	Related Genes	Homo sapiens
KEGG_PATHWAY	Tryptophan metabolism, Methane metabolism, Amyotrophic lateral sclerosis (ALS),		
CCR6, Ccnl2	cyclin L2; chemokine (C-C motif) receptor 6	Related Genes	Homo sapiens
KEGG_PATHWAY	Cytokine-cytokine receptor interaction, Chemokine signaling pathway,		

Fig 24: functional annotation of table .

NPY	neuropeptide Y	Related Genes	Homo sapiens
KEGG_PATHWAY	Adipocytokine signaling pathway,		
Slc1a2	solute carrier family 1 (glial high affinity glutamate transporter), member 2	Related Genes	Homo sapiens
KEGG_PATHWAY	Amyotrophic lateral sclerosis (ALS),		
Sod2	superoxide dismutase 2, mitochondrial	Related Genes	Homo sapiens
KEGG_PATHWAY	Huntington's disease,		
TSLP	thymic stromal lymphopoietin	Related Genes	Homo sapiens
KEGG_PATHWAY	Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway,		
TLR1	toll-like receptor 1	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
TLR2	toll-like receptor 2	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
TLR3	toll-like receptor 3	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
TLR4	toll-like receptor 4	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway, Pathogenic Escherichia coli infection,		
TLR6	toll-like receptor 6	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
TLR7	toll-like receptor 7	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
TICAM1	toll-like receptor adaptor molecule 1	Related Genes	Homo sapiens
KEGG_PATHWAY	Toll-like receptor signaling pathway,		
Tgfbr2	transforming growth factor, beta receptor II (70/80kDa)	Related Genes	Homo sapiens
KEGG_PATHWAY	MAPK signaling pathway, Cytokine-cytokine receptor interaction, Endocytosis, TGF-beta signaling pathway, Adherens junction, Pathways in cancer, Colorectal cancer, Pancreatic cancer, Chronic myeloid leukemia,		
tubb3, MC1R	tubulin, beta 3; melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	Related Genes	Homo sapiens
KEGG_PATHWAY	Neuroactive ligand-receptor interaction, Gap junction, Melanogenesis, Pathogenic Escherichia coli infection,		
TYRL, tyr	tyrosinase-like (pseudogene); tyrosinase (oculocutaneous albinism IA)	Related Genes	Homo sapiens
KEGG_PATHWAY	Tyrosine metabolism, Riboflavin metabolism, Melanogenesis,		

Fig25: functional annotation of table .

Diseases related to genes found out shown below. Maximum diseases is found out to be related with PTPN22. Here in the table given above functional annotation of 40 genes have been found out with their related genes and pathway.

40 record(s) Download File

Mthfr	5,10-methylenetetrahydrofolate reductase (NADPH)	Related Genes	Homo sapiens
OMIM_DISEASE	Cleft lip/palate, susceptibility to, Homocystinuria due to MTHFR deficiency, Neural tube defects, susceptiblity to, Schizophrenia, susceptibility to, Thromboembolism, susceptibility to, Vascular disease, susceptiblity to,		
Bach2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	Related Genes	Homo sapiens
OMIM_DISEASE	Follow up analysis of genome-wide association data identifies novel loci for type 1 diabetes, Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci,		
CD44	CD44 molecule (Indian blood group)	Related Genes	Homo sapiens
OMIM_DISEASE	Blood group, Indian system,		
FANCA	Fanconi anemia, complementation group A	Related Genes	Homo sapiens
OMIM_DISEASE	Fanconi anemia, complementation group A,		
IKZF4	IKAROS family zinc finger 4 (Eos)	Related Genes	Homo sapiens
OMIM_DISEASE	A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study,		
KITLG	KIT ligand	Related Genes	Homo sapiens
OMIM_DISEASE	Genetic determinants of hair, eye and skin pigmentation in Europeans, Skin/hair/eye pigmentation 7, blond/brown hair, [Skin/hair/eye pigmentation 7, blond/brown hair],		
lpp	LIM domain containing preferred translocation partner in lipoma	Related Genes	Homo sapiens
OMIM_DISEASE	Leukemia, acute myeloid, Lipoma, Newly identified genetic risk variants for celiac disease related to the immune response,		
Nlrp1	NLR family, pyrin domain containing 1	Related Genes	Homo sapiens
OMIM_DISEASE	Systemic lupus erythematosus, vitiligo-related, susceptibility to, 1, Vitiligo-associated multiple autoimmune disease susceptibility 1,		
RPGRIP1L	RPGRIP1-like	Related Genes	Homo sapiens
OMIM_DISEASE	Joubert syndrome 7, Meckel syndrome, type 5,		
SH2B3	SH2B adaptor protein 3	Related Genes	Homo sapiens
OMIM_DISEASE	Celiac disease, susceptiblity to, 13, Diabetes mellitus, insulin-dependent, susceptibility to, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, Newly identified genetic risk variants for celiac disease related to the immune response,		
Xbp1	X-box binding protein 1	Related Genes	Homo sapiens
OMIM_DISEASE	Bipolar disorder, susceptibility to, Major affective disorder-7, susceptibility to,		
ATXN2	ataxin 2	Related Genes	Homo sapiens
OMIM_DISEASE	Newly identified genetic risk variants for celiac disease related to the immune response, Spinocerebellar ataxia-2,		
BMP6	bone morphogenetic protein 6	Related Genes	Homo sapiens
OMIM_DISEASE	Many sequence variants affecting diversity of adult human height,		
BTNL2	butyrophilin-like 2 (MHC class II associated)	Related Genes	Homo sapiens
OMIM_DISEASE	Sarcoidosis, susceptibility to, Sarcoidosis, susceptibility to, 2, Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility,		
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	Related Genes	Homo sapiens
OMIM_DISEASE	Breast cancer, lobular, Cleft lip with or without cleft palate, with gastric cancer, familial diffuse, Endometrial carcinoma, Gastric cancer, familial diffuse, Listeria monocytogenes, susceptibility to, Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer, Ovarian carcinoma,		
cat	catalase	Related Genes	Homo sapiens
OMIM_DISEASE	Acatalasemia,		
CCR6, Ccnl2	cyclin L2; chemokine (C-C motif) receptor 6	Related Genes	Homo sapiens
OMIM_DISEASE	Genome-wide assocation defines more than 30 distinct susceptibility loci for Crohn's disease,		
CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	Related Genes	Homo sapiens

Fig 26 : figure is showing the genes and their related diseases

Functional category.

Three functional categories have been found out. These are :

1) Cog ontology : This ontology includes involvement of genes in various cellular processes.

2) Sp-pir-keywords : This shows the functional annotation table with related genes for each node

3) Up-seq-feature: This shows types of enzyme it produce and the factors involved in each process.

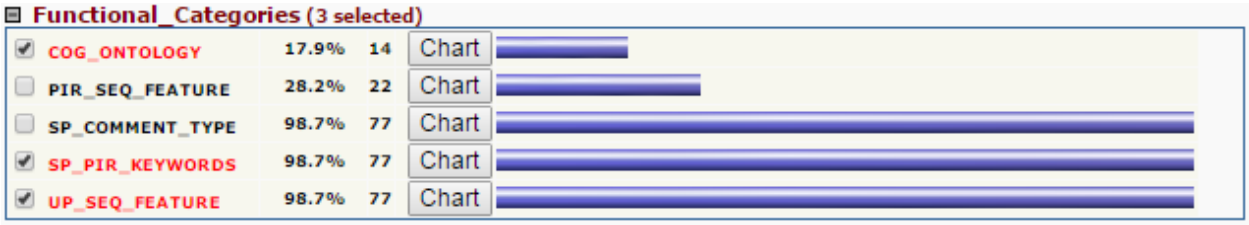
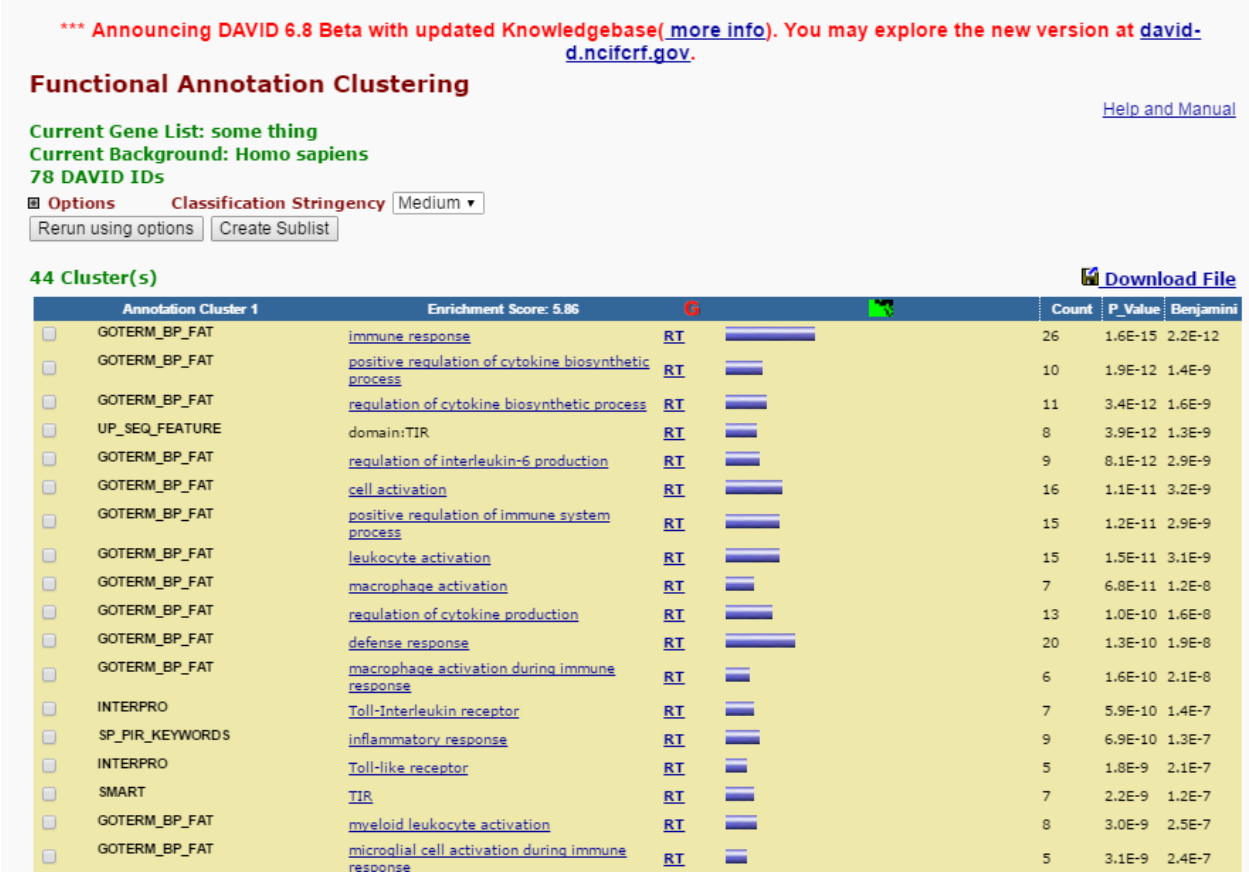


Fig 27: Result of functional categories.

Functional annotation clustering.

44 clusters are present based on protein classification are present in the data of the gene as given below . It contain enrichment score , p-value , count etc.



*** Announcing DAVID 6.8 Beta with updated Knowledgebase(more info). You may explore the new version at david-d.ncifcrf.gov.

Functional Annotation Clustering

Help and Manual

Current Gene List: some thing
Current Background: Homo sapiens
78 DAVID IDs

Options Classification Stringency Medium

Rerun using options | Create Sublist

44 Cluster(s)

Download File

Annotation Cluster 1	Enrichment Score: 5.86	G		Count	P_Value	Benjamini
GOTERM_BP_FAT	immune response	RT		26	1.6E-15	2.2E-12
GOTERM_BP_FAT	positive regulation of cytokine biosynthetic process	RT		10	1.9E-12	1.4E-9
GOTERM_BP_FAT	regulation of cytokine biosynthetic process	RT		11	3.4E-12	1.6E-9
UP_SEQ_FEATURE	domain:TIR	RT		8	3.9E-12	1.3E-9
GOTERM_BP_FAT	regulation of interleukin-6 production	RT		9	8.1E-12	2.9E-9
GOTERM_BP_FAT	cell activation	RT		16	1.1E-11	3.2E-9
GOTERM_BP_FAT	positive regulation of immune system process	RT		15	1.2E-11	2.9E-9
GOTERM_BP_FAT	leukocyte activation	RT		15	1.5E-11	3.1E-9
GOTERM_BP_FAT	macrophage activation	RT		7	6.8E-11	1.2E-8
GOTERM_BP_FAT	regulation of cytokine production	RT		13	1.0E-10	1.6E-8
GOTERM_BP_FAT	defense response	RT		20	1.3E-10	1.9E-8
GOTERM_BP_FAT	macrophage activation during immune response	RT		6	1.6E-10	2.1E-8
INTERPRO	Toll-Interleukin receptor	RT		7	5.9E-10	1.4E-7
SP_PIR_KEYWORDS	inflammatory response	RT		9	6.9E-10	1.3E-7
INTERPRO	Toll-like receptor	RT		5	1.8E-9	2.1E-7
SMART	TIR	RT		7	2.2E-9	1.2E-7
GOTERM_BP_FAT	myeloid leukocyte activation	RT		8	3.0E-9	2.5E-7
GOTERM_BP_FAT	microglial cell activation during immune response	RT		5	3.1E-9	2.4E-7

Fig 28 : Functional annotation clustering.

KEGG pathway analysis

Kegg pathway database tool is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for: Metabolism Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan .Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure, Genetic Information Processing , Environmental Information Processing , Cellular Processes , Organismal Systems , Human Diseases and also on the structure relationships (KEGG drug structure maps) in Drug Development. On observing the genes find in kegg pathway of vitiligo is almost similar to the genes find out using various database tool. This validates result obtained after collecting data from various sources.

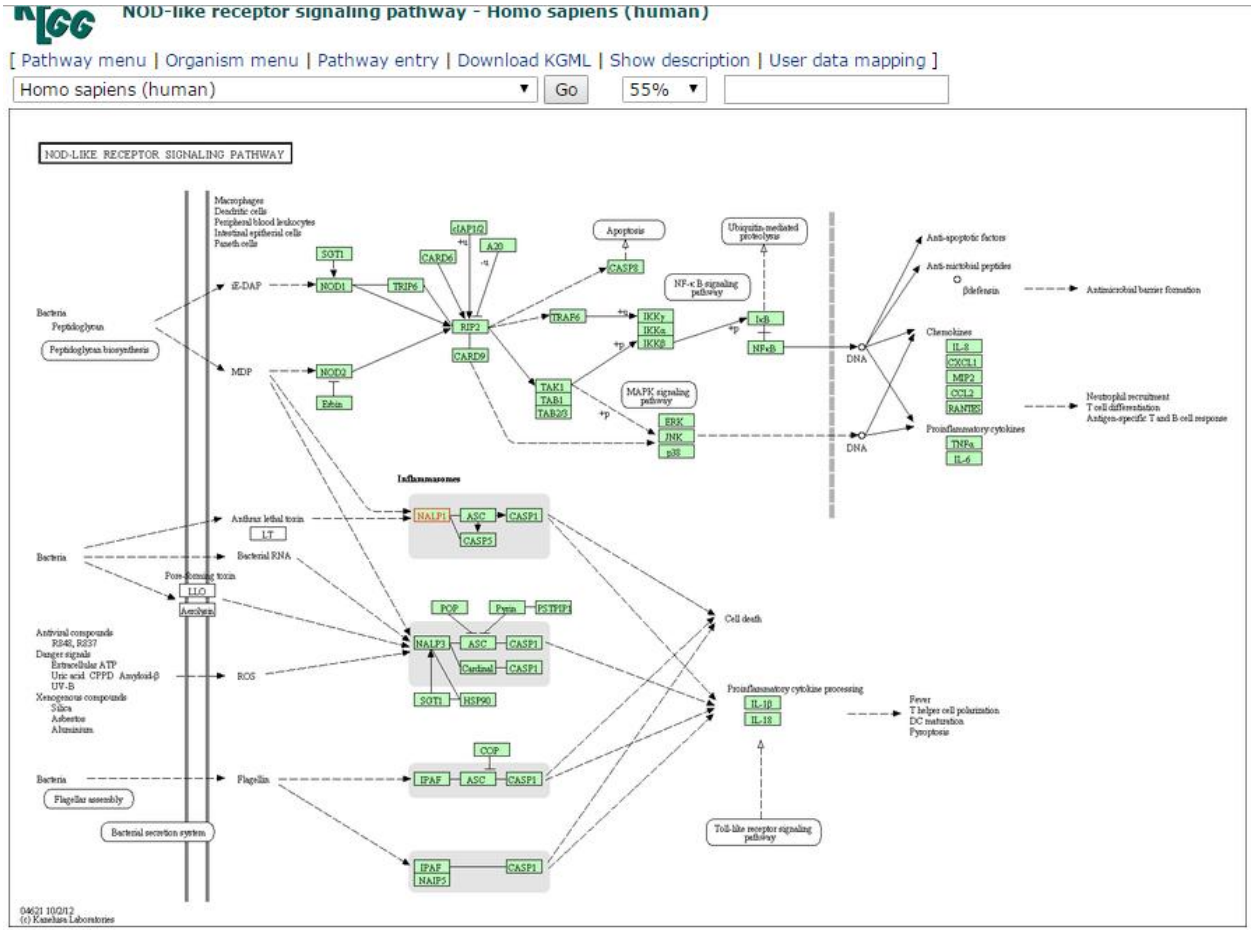


Fig 29: Genes and the protein involved in vitiligo found in KEGG pathway.(04621 10/2/12 (c) kanehisa laboratories.)

Discussion and conclusion

During all my work It is found out that there are lots of pathways having genes common in vitilgo and many more diseases shows that genes responsible for vitiligo is also responsible for many more diseases which is shown in result of DAVID bioinformatics tool. It is found out that there are approximately 2600 pathways in which the obtained genes are present. During this project it is established that most of the genes are present on chromosome no. 2, 5 , 6, 10,11,12,14,15,16,17,22 and it has been find out that this disease is mostly found in han Chinese, indian sub continent , Caucasian, Korean , north American , African, Estonian population. During this project It is found out that some of the gene which could produce hub protein which is a key finding in the project are **FOXP3 , MHC , CTLA4 , ICOS.** Kappa score of **0.4** shows good confidence ratio in finding out connectivity and coorelationship between nodes. Using p-value as a source of interaction between genes It comes to the knowledge that there is no clustering but It is amazing to found that there are 44 clusters in functional annotation clustering when I don`t used p-value as interaction source. This shows that for clustering of genes to occour probability value (p-value) should not be defined. This is a very authentic and genuine study in relation to vitiligo as it provide some very authentic unanswered question which has been discussed in the “result and discussion” session. Apart from this it is a work which is a step forward in moving one step forward in understanding pathway and machinery of vitiligo to spread from one body part to other body part.

Refrences

1) Akbas H1, Dertlioglu SB2, Dilmec F1, Atay AE3 (feb 2014) . Lack of Association between PTPN22 Gene +1858 C>T Polymorphism and Susceptibility to Generalized Vitiligo in a Turkish Population.

2) Akbas H1, Dertlioglu SB, Dilmec F, Balkan M (may-june 2013) . No association between catalase (CAT) gene polymorphisms and susceptibility to vitiligo in a Turkish population.

3) Alzolibani AA1, Al Robaee A1, Al-Shobaili H1, Al-Saif F2, Al-Mekhadab E2, Settin AA3. (June 2014) Association of CYP2C9 Genetic Variants with Vitiligo.

4) Chen JX1, Shi Q, Wang XW, Guo S, Dai W, Li K, Song P, Wei C, Wang G, Li CY, Gao TW (may 2014) Genetic polymorphisms in the methylenetetrahydrofolate reductase gene (MTHFR) and risk of vitiligo in Han Chinese populations: a genotype-phenotype correlation study.

5) Cheong KA1, Kim NH, Noh M, Lee AY (may 2013) . Three new single nucleotide polymorphisms identified by a genome-wide association study in Korean patients with vitiligo.

6) Dwivedi M1, Laddha NC, Mansuri MS, Marfatia YS, Begum R (nov 2013) . Association of NLRP1 genetic variants and mRNA overexpression with generalized vitiligo and disease activity in a Gujarat population.

7) Dwivedi M1, Laddha NC, Shah K, Shah BJ, Begum R (nov 2013). Involvement of interferon-gamma genetic variants and intercellular adhesion molecule-1 in onset and progression of generalized vitiligo.

8) Imran M1, Laddha NC, Dwivedi M, Mansuri MS, Singh J, Rani R, Gokhale RS, Sharma VK, Marfatia YS, Begum R (aug-2012). Interleukin-4 genetic variants correlate with its transcript and protein levels in patients with vitiligo.

9)Jin Y1, Birlea SA, Fain PR, Mailloux CM, Riccardi SL, Gowan K, Holland PJ, Bennett DC, Wallace MR, McCormack WT, Kemp EH, Gawkrodger DJ, Weetman AP, Picardo M, Leone G,

Taïeb A, Jouary T, Ezzedine K, van Geel N, Lambert J, Overbeck A, Spritz RA(july 2010) . Common variants in FOXP1 are associated with generalized vitiligo.

10) Jin Y1, Birlea SA, Fain PR, Ferrara TM, Ben S, Riccardi SL, Cole JB, Gowan K, Holland PJ, Bennett DC, Luiten RM, Wolkerstorfer A, van der Veen JP, Hartmann A, Eichner S, Schuler G, van Geel N, Lambert J, Kemp EH, Gawkrodger DJ, Weetman AP, Taïeb A, Jouary T, Ezzedine K, Wallace MR, McCormack WT, Picardo M, Leone G, Overbeck A, Silverberg NB, Spritz RA (6-may 2012).Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo.

11) Kim SK1, Hong MS, Shin MK, Uhm YK, Chung JH, Lee MH(no-2011) . Promoter polymorphisms of the HLA-G gene, but not the HLA-E and HLA-F genes, is associated with non-segmental vitiligo patients in the Korean population.

12) Laddha NC1, Dwivedi M1, Mansuri MS1, Singh M1, Patel HH1, Agarwal N1, Shah AM1, Begum R1(15 sep 2014) .Association of neuropeptide Y (NPY), interleukin-1B (IL1B) genetic variants and correlation of IL1B transcript levels with vitiligo susceptibility.

13) Laddha NC1, Dwivedi M, Gani AR, Shajil EM, Begum R (dec 2013) .Involvement of superoxide dismutase isoenzymes and their genetic variants in progression of and higher susceptibility to vitiligo.

14) Li S1, Yao W, Pan Q, Tang X, Zhao S, Wang W, Zhu Z, Gao J, Sheng Y, Zhou F, Zheng X, Zuo X, Sun L, Zhang A.(2015) Institute of Dermatology & Department of Dermatology, The First Affiliated Hospital of Anhui Medical University, Hefei, 230032, China. Association analysis revealed one susceptibility locus for vitiligo with immune-related diseases in the Chinese Han population.

15) Ozel Turkcu U1, Tekin NS2, Edgunlu TG3, Karakas SC4, Oner S5.(20 december 2013) The association of Foxo3a gene polymorphisms with serum Foxo3a levels and oxidative stress markers in vitiligo patients.

16) Shin MK1, Im SH, Park HJ, Kim SK, Yim SV, Chung JH, Lee MH. (nov 2011) .Association study between polymorphisms of CD28, CTLA4 and ICOS and non-segmental vitiligo in a Korean population.

17) Shin MK1, Uhm YK, Lee JH, Kim SK, Chung JH, Lee MH (july –aug 2012) . Association between CDK5RAP1 polymorphisms and susceptibility to vitiligo in the Korean population.

18) Song P1, Wang XW, Li HX, Li K, Liu L, Wei C, Jian Z, Yi XL, Li Q, Wang G, Li CY, Gao TW (sep 2013) . Association between FOXP3 polymorphisms and vitiligo in a Han Chinese population.

19) Tang J1, Liu JL, Zhang C, Hu da Y, He SM, Zuo XB, Wang PG, Sun LD, Zhang XJ, Yang S (june 2013) . The association between a single nucleotide polymorphism rs11966200 in MHC region and clinical features of generalized vitiligo in Chinese Han population.

20) Tang XF1, Zhang Z, Hu DY, Xu AE, Zhou HS, Sun LD, Gao M, Gao TW, Gao XH, Chen HD, Xie HF, Tu CX, Hao F, Wu RN, Zhang FR, Liang L, Pu XM, Zhang JZ, Han JW, Pan GP, Wu JQ, Li K, Su MW, Du WD, Zhang WJ, Liu JJ, Xiang LH, Yang S, Zhou YW, Zhang XJ (feb – 2013) .Association analyses identify three susceptibility Loci for vitiligo in the Chinese Han population.