DECIPHERING PATHWAYS ASSOCIATED WITH TUBERCULOSIS USING GENE EXPRESSION ANALYSIS

M.Tech (Industrial Biotechnology)

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

INDUSTRIAL BIOTECHNOLOGY

Submitted by:

SHASHI BALA YADAV

2k20/IBT/08

Under the supervision of

Prof. YASHA HASIJA

Shashi Bala Yadav

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

MAY 2022

2022

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi – 110042

CANDIDATE'S DECLARATION

I, **Shashi Bala Yadav**, Roll No. **2k20/IBT/08**, student of **M.Tech in Industrial Biotechnology**, hereby declare that the Project Dissertation titled **"Deciphering pathways associated with tuberculosis using gene expression analysis"** which is submitted by me to the Department of Biotechnology, Delhi Technical University Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **SHASHI BALA YADAV**

Date:

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi – 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **"Deciphering pathways associated with tuberculosis using gene expression analysis"** which is submitted by **Shashi Bala Yadav**, Roll No. **2k20/IBT/08**, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is record of the project work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

PROFESSOR YASHA HASIJA
SUPERVISOR
Professor
Department of Biotechnology
Delhi Technological University

PROFESSOR PRAVIR KUMAR
HEAD OF DEPARTMENT (HOD)
Department of biotechnology
Delhi Technological University

Date:

Place: Delhi

ABSTRACT

Every year there are almost 2-3 million deaths by latent tuberculosis persists in over a billion individuals worldwide. It is being very clear that TB is spread through the air when people with lung TB cough, sneeze, or spit. A person needs to inhale only a few germs to become infected. As a result, there is an essential need for an innovative approach to tuberculosis treatment. During chronic infection with M. tuberculosis, or early innate immune response the phagocytes initiate internalization of M. tuberculosis through the pathogenic antigen. Phytochemicals,which are plant-derived metabolites that act as anti-mycobacterial are used in a wide variety of medical applications. In this work, we applied computational techniques to evaluate the therapeutic potential of diverse plant phytochemicals that can be used against diabetes, as well as the control of differentially expressed genes via molecular pathways and biological processes which will aid in tuberculosis management. We analyzed a RNA seq dataset for diabetes taken from the GEO database and further processed it as per requirements into an usable report. A total of 379 genes were identified out of which 16 were getting upregulated and 363 downregulated genes. These genes were then matched with the phytochemicals from the plants that are known to have an effect for diabetes. Later on, gene enrichment analysis revealed that the downregulated DEGs were mainly enriched in the biological processes like regulation of apoptotic signalling, mitochondrial depolarisation and in Angiotensin-Activated Signalling. Upregulated DEGs were basically related to the biological process like cellular responsesto organic cyclic compounds. Finally, hub genes and hub modules were identified that could show a potential significance in the Diabetes

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to our research guide **Professor Yasha Hasija, Department of Biotechnology, Delhi Technological University** for her unstinted inspiration, invaluable guidance, encouragement, keen interest, good wishes and valuable suggestions throughout my entire research tenure.

I express my kind regards and gratitude to **Professor Pravir Kumar, Head of Department, Department of Biotechnology, Delhi Technological University** and all the faculty members for helping in my project.

I am thankful to **Raj Kumar Chakraborty sir** and all my classmates for giving me moral boost and making my hopes alive with energy and enthusiasm to carry out the present work and helping hand at every step. My sincere thanks to my parents and family members for encouragement and moral support I received during the tenure of study.

Place: Delhi **SHASHI BALA YADAV**

Date:

CONTENTS

CONTENT	PAGE NO.
CANDIDATE'S DECLARATION	II
CERTIFICATE	III
ABSTRACT	IV
ACKNOWLEDGEMENT	V
CONTENTS	VI
LIST O FIGURES	VII
LIST OF TABLES	VIII
1. INTRODUCTION	1 - 3
2. LITERARY REVIEW	4 - 12
3. MATERIAL & METHODS	13- 18
4. RESULTS	19 - 27
5. DISCUSSION	28 - 30
6. CONCLUSION	31
7. REFERENCE	32-36

LIST OF FIGURES

S.No.	CONTENT	PAGE NO.
1	Inhalation of M. tuberculosis bacilli into the lung formation of granuloma during TB infection	8
2	Processing of dataset obtained from GEO - Volcano plot showing upregulated	20
3	Cytoscape network for medicinal plant (onion) and its associated phytochemicals	21
4	PPI network with the matched DEGs	24
5A - 5D	Top Gene Ontology (GO) enrichment results along with associated genes in biological process, molecular function, cellular component and the wikipathway of the phytochemical associated matched genes	26-27

LIST OF TABLES

S.No.	CONTENT	PAGE NO.
1	Classification of mycobacteria	4
2	Tuberculosis treatment regimen.	6
3A & 3B	Brief information on the selected RNA-sequenced dataset	13-16
4	List of phytochemicals of onion plant showing matched genes	21-23

CHAPTER 1

INTRODUCTION

Chronic disorders (CDC) are the set of diseases that are persistent or long-lasting minimum three months and its effects are gradually seen. Tuberculosis, Diabetes, Cancer, Stroke, Arthritis are some well-known chronic disorders.

TB is one of the oldest recorded catastrophes caused by slow-growing pathogenic bacteria M. tuberculosis and it most often affects the lower respiratory tract. Every year there are approximately 2-3 million deaths by latent tuberculosis persists in over a billion individuals worldwide. It is being very clear that TB is spread through the air when people with lung tuberculosis cough, sneeze, or spit.

LTBI is defined as a clinical conditioning which a host is chronically infected with M. tuberculosis but without evidence of clinically manifested active Tuberculosis disease. The most commonly used diagnostic tool for tuberculosis is a simple skin test i.e. known as Mantoux tuberculin skin test (TST), though blood tests are becoming more common place. Additional tests are required to confirm TB disease.

Drug therapies for tubercuosis have not been changed significantly in the past several decades although this current treatment regimen is very long and very complicated. The currently established regimen requires various drugs to be taken simultaneously; by which there is chance of increasing the patient's risk of harmful drug interactions. All of these factors discussed contributes to suboptimal patient adherence to the current treatment which leads to further propagation of infectious drug-resistant strains of TB.

Gene expression is an organic-phenomena by which a gene is regulated within a cell to form RNA and proteins. Genetic variation further leads to a disease due to faulty regulation of gene expression. With the advancement of technology and to decrease the

challenge of analysing large gene expression datasets, bioinformaticians had developed various new solutions. Metabolite profiling and massiveprogress in high throughput sequencing gave rise to a new domain of 'phytochemical genomics'. Phytochemicals are plant metabolites that help to modulate coding and non-coding RNA gene expression which helps in normal functioning of the metabolic pathways and also for treating chronic diseases caused by oxidative stress. Thus, phytotherapies are gaining attention nowadays for treating diabetes and finding new drug model. Phytochemicals against tuberculosis taken from medicinal plants provides identical waysfor developing functional foods and anti-mycobacterial drugs.

For better understanding of the corelation between genes with the pathways for various diseases researchers took the help of computation tools like Enrichr for pathway enrichment analysis and Cytoscape for functional analysis. It has been reported that plants like onion, ginger, tea and phytochemicals (i.e. curcumin and resveratrol) with an impact on tuberculosis.

In this study, we have illustrated the role of plant phytochemicals whose genes are regulated depending on the pathway for T2D. Further, we have explored a number of computational databases like:

(1) Gene Expression Omnibus (GEO) which is a repositoryfor RNA-seq, microarray, chip data

(2) BioJupies a web application for generation of customized notebooks of raw dataset obtained from GEO.

(3) IMPPAT a manually curated largest database which provide information about the

medicinal plants and its related phytochemicals along with their canonical smiles that help in figuring out the chemical structure.

(4) STRING database for predicting the protein-protein interaction

(5) Enrichr for analysis of the cellular component, molecular functionality and enriched biological pathways.

CHAPTER 2

LITERATURE REVIEW

2.1 Tuberculosis (TB)

TB is one of the oldest recorded catastrophes, first reported by Robert Koch in 1882, caused by slow-growing pathogenic bacteria M. tuberculosis (see Table 1) and it most often affects the lower respiratory tract. Every year there are almost 2-3 million deaths by latent tuberculosis infection persists in over a billion individuals worldwide. Tuberculosis can be spread through the air when the people with lung Tuberculosis cough, sneeze, or spit in the environment. A person becomes infected when only a few germs are inhaled by them. Tuberculosis pateints experience the following symptoms like fever, cough, and weight loss, and the diagnosis of TB can usually be confirmed with culture, sputum smear, and molecular tests. Often, these symptoms will be mild for many months after the infection [1].

S.No.	**Mycobacteria belongs to**	
1.	Kingdom	Bacteria
2.	Phylum	Actinobacteria
3.	Order	Actinomycetales
4.	Family	Mycobacteriaceae
5.	Genus	Mycobacterium

Table 1: - Classification of mycobacteria

2.2 The genome of M. *tuberculosis*

The total genome architecture of the virulent strain of Mycobacterium *tuberculosis*, which is H37Rv was completely sequenced and published in 1998. It presents a sequence of 4.4 X106 bps and encode 4,000 genes and high guanine plus cytosine (G+C) content (65.5%) [2].

2.3 Classification of TB

Although TB disease can be viewed as a dynamic continuum from M. tuberculosis infection to active infectious disease, patients are categorized as having either latent TB infection (LTBI) or active TB disease for simplicity in clinical and public health settings. LTBI is defined as a clinical conditioning which a host is chronically infected with M. tuberculosis but without evidence of clinically manifested active TB disease [4,5].

2.4 Diagnosis and Treatment of Tuberculosis

The most commonly used diagnostic tool for tuberculosis is a simple skin test i.e. known as Mantoux tuberculin skin test (TST), though blood tests are becoming more common place. Additional tests are required to confirm TB disease. A complete diagnosis process or medical assessment for TB must include a chest X-ray, a medical history, and a culture based examination and also, acid fast staining and PCR, or True NAT/Mtb expert are PCR based assays [4]. Drug therapies for tubercuosis have not been changed significantly in the past several decades although this current treatment regimen is very long and very complicated. The currently established regimen requires various drugs to be taken simultaneously; by which there is chance of increasing the patient's risk of harmful drug interactions. All of these factors discussed contributes to suboptimal patient adherence to the current treatment which leads to further propagation of infectious drug-resistant strains of TB [7].

	Initial Phase			Continuation Phase			
Regimen	Drugs	Interval	Doses	Drugs	Interval	Doses	Total
1	INH	7 Days/Week	56 doses	INH	7 Days/Week	126 doses	182-130 doses, 26 weeks
	RIF	5 Days/Week	40 doses	RIF	7 Days/Week	126 doses	
	PZA	5 Days/Week	40 doses				
	EMB						
2	INH	7 Days/Week	56 doses	INH	7 Days/Week	217 doses	273-195 doses, 39 weeks
	RIF	5 Days/Week	40 doses	RIF	7 Days/Week	217 doses	
	EMB	5 Days/Week	40 doses				

Table 2:- Tuberculosis treatment regimen.

2.5 Global TB Data

In 2019, based on the surveillance and survey data WHO published Global TB report, worldwide every year approx. 10 million people fall ill with Tuberculosis worldwide. Tuberculosis is included in top 10 causes of death of human by a single infectious agent. These diseases can affect anyone anywhere, but in 90% of cases, TB is mostly developed in adults. The is ratio of male and female infected with TB is 2:1. Although TB is preventable curable and preventable disease, there are 1.5 million people die from TB every year which makes TB the world's top infectious killer. About half of all people with TB can be found in 8 countries: China, Bangladesh, Indonesia, India, Pakistan, Nigeria, South Africa and Philippines (Global Tuberculosis report). According to the WHO, there are 484 000 new cases of tuberculosis with resistance to rifampicin (which is the most effective first-line drug of tuberculosis) of which 78% pateints had MDR-TB.

2.7 Intracellular environment and adaptation

During chronic infection with M. tuberculosis, or early innate immune response the phagocytes

initiate internalization of M. tuberculosis through the pathogenic antigen interacts with certain host receptors, mainly the Toll-like receptors (TLR)-2 present in the phagosome, hence effecting lipid body formation in macrophages. Conversion of normal macrophages into foamy ones happens due to an imbalance in the low-density lipoprotein (LDL) fraction [16]. There are many types of LDL such as phospholipids and triacylglycerol which get metabolized whereas cholesterol undergoes esterification and gets retained in the macrophages and later deported into lipid droplets. Accumulation of foamy macrophages in granulomas during infection was proposed to be due to the action of mycolic acid [17], but later refuted by studies that showed that Mtb infection on its own did not indice foamy macrophage formation [13–15]. At later stages of infection, the bacteria are contained in phagosomes and are located near lipid bodies [17]. Additionally, the bacteria are known to utilize fatty acid release from host triacylglycerol and incorporate into its triacylglycerol pools, primarily by the action of bacterial triacylglycerol synthase. M. tuberculosis is known to tolerate the lipid-rich adipocytes thus suggesting its ability to survive by manipulation of host lipid metabolism[18,19].

2.8 The cellular immune response to M. tuberculosis

As I have mentioned tuberculosis disease is spread by airborne droplet, reside inside the macrophage and early host response to M. tuberculosis infection is characterized by alveolar macrophage and neutrophils. Therefore it is important to understand how the immune system is affected due to infection? What role do immune cells play during infection?.

2.9 Regulation of the immune response during M. tuberculosis infection.

Once the entry of M. tuberculosis inside the lung recognition by alveolar macrophages and dendritic cells is followed by an inflammatory response. The alveolar macrophage, specific regulator pathways that normally serve to limit host-induced immune pathology may promote pathogen survival inside the host. Two such regulators include cytokine IL-10 and regulatory T cells). It is

more suppressing that IL-10 produce by several immune cells in the host such as neutrophils, macrophages, B cells, DCs, and T cells. The stimulation of IL-10 during infection act as an immunosuppressive cytokine and lead to blocking chemotactic factors that control dendritic cells trafficking to the demanding lymph nodes. Besides, it is also involved in inhibition of macrophage effector functions and delays production of the cytokines IFN, IFN-γ and IL-17 by CD4+ T cells in the lung, with reduced bacterial killing and decreased secretion of cytokines/chemokines.

Gene expression is a process by which information from a gene is utilised to build a protein. It is considered as an efficient way to predict the mechanisms of toxicity. It continuously compares the RNA expressions of several genes. Differential expression ingenes is basically due to genetic alterations/ variations. When a mutation is seen in a protein that is crucial in body functioning then diseases are caused [16]. It is important inunderstanding the biological differences between a healthy and a diseased person. DGE is associated with methylation of DNA [16]. As mentioned before techniques like high throughput sequencing (RNA-Seq), cellular microarray chip methods are mainly used to scan the gene expressions of large number of test samples. This advances in technologies allowed the biomedical sector to undergo more research on cell therapeutics and drug discovery. RNA-Seq technique is more powerful than microarray for transcriptome analysis as it helps in finding more DEGs and gave a wider quantitative range of expression changes. RNA-Seq data helps to identify non-coding DEGs, new transcripts like fusion genes (that are formed due to chromosome rearrangement), also detect SNPs and compare between a particular data and reference genome (1000 genome project) [17].A different method used is ‘exome sequencing’ where only particular nucleotide changesin the coding regions that results to a specific phenotype. Recently, research finds this technique more useful. edgeR, DESeq2, R package are some well-known software used for differential gene expression analysis [18].

From the classical experiments it was concluded that proteins are important for biological functions and also for predicting the phenotypes. And with the advancement of science, it was found that proteins naturally are not functional rather when interacted with different molecules like DNA, RNA they help in stimulating signalling pathways, cellular components [19]. Thus, analysis of protein interaction become a crucial part for bioinformaticians. Protein-protein interaction data can be utilised is several studies relating to identifying new functionality of a protein, recognizing the genotype and phenotype associations, drug development in biomedical sector. Proteins interactions helps in differentiating control and diseased samples molecular basis and cellular functionality. Information about protein interaction can be gathered by literature mining and also from high-throughput methods like yeast-2-hybrid techniques (Y2H). Among allthe proteins used for predicting the network between them some are called 'hub proteins'as they show high number of interactions [20]. Some commonly available software for determining protein interactions: Cytoscape, ShinyGO, QuickGO. While using the software some standard indices are used like average degree(K), clustering coefficient(C), average path length (L), diameter(D). The protein networks help in identifying the pathways related to a disease. Some recent findings says that almost 39,000 protein interactions have been identified in humans, disease genes have a tendency to code for the highly interacted proteins and also try to cluster together in the network positions, proteins of same kind of phenotype are mainly interconnected [21]. In a study it has beenfound that variation on genes of linked proteins results in same disease as they share a similar functional relation. Thus, PPI also help in prioritize the genes for identifying the background genetic role in a disease [22]. Once the networks are clearly understood, drugdesigning becomes easy for the researchers as they can easily identify the potential drugtargets. If a drug target is found to be a hub gene, then its inhibition can affect various other interaction thus, making it not a suitable target point. So, a gene that is less

interconnected should always be chosen a potential drug targeting point.

For examining modification in the genetic expressions, its related pathways are studied. Gene and its pathway enrichment analysis is another essential part of in-silico studies to interpret datasets containing candidate genes for their biological process, cellular component [23]. Some tools for doing gene set enrichment are: Enrichr tool that is for mammalian gene sets, accessed through API and helps in visualizing the interactions in the form of charts and graphs, AmiGO2 developed by Gene Ontology group, Blast2GO platform for determining functional annotation and genomic datasets.

CHAPTER 3

MATERIAL AND METHODS

3.1 Preparation of Dataset

Publicly available RNA-sequenced dataset was obtained from NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) database [24]. All the tissue samples were obtained from whole blood of infected and non-infected patients of tuberculosis. Basic information about the dataset is shown in Table 3A and 3B.

Table 1: Brief information on the selected dataset.

Tissue Origin	Geo Accession Number	Sequencing Technique Used	Sample Size	
			Control	Patient
Whole Blood	GSE107993 [25]	RNA Sequencing OR High Throughout Sequencing	69	69

(A)

Sample_geo_accession	Sample Title	Gender	Group
GSM2886136	Leicester_non_progressor_longitudnal_only_Sample2	M	Control
GSM2886137	Leicester_non_progressor_longitudnal_only_Sample3	M	Control
GSM2886138	Leicester_non_progressor_longitudnal_only_Sample4	M	Control
GSM2886139	Leicester_non_progressor_longitudnal_only_Sample5	M	Control
GSM2886140	Leicester_non_progressor_longitudnal_only_Sample6	M	Control
GSM2886141	Leicester_non_progressor_longitudnal_only_Sample7	M	Control
GSM2886142	Leicester_non_progressor_longitudnal_only_Sample10	M	Control
GSM2886143	Leicester_non_progressor_longitudnal_only_Sample11	M	Control
GSM2886144	Leicester_non_progressor_longitudnal_only_Sample12	M	Control
GSM2886145	Leicester_non_progressor_longitudnal_only_Sample13	M	Control
GSM2886146	Leicester_non_progressor_longitudnal_only_Sample14	M	Control
GSM2886147	Leicester_non_progressor_longitudnal_only_Sample15	M	Control
GSM2886148	Leicester_non_progressor_longitudnal_only_Sample16	M	Control
GSM2886149	Leicester_non_progressor_longitudnal_only_Sample17	M	Control
GSM2886150	Leicester_non_progressor_longitudnal_only_Sample18	F	Control
GSM2886151	Leicester_non_progressor_longitudnal_only_Sample19	F	Control
GSM2886152	Leicester_non_progressor_longitudnal_only_Sample20	F	Control
GSM2886153	Leicester_non_progressor_longitudnal_only_Sample21	F	Control
GSM2886154	Leicester_non_progressor_longitudnal_only_Sample22	M	LTBI

GSM2886155	Leicester_non_progressor_longitudnal_only_Sample23	M	LTBI
GSM2886156	Leicester_non_progressor_longitudnal_only_Sample24	M	LTBI
GSM2886157	Leicester_non_progressor_longitudnal_only_Sample25	M	LTBI
GSM2886158	Leicester_non_progressor_longitudnal_only_Sample26	M	LTBI
GSM2886159	Leicester_non_progressor_longitudnal_only_Sample27	M	LTBI
GSM2886160	Leicester_non_progressor_longitudnal_only_Sample28	M	LTBI
GSM2886161	Leicester_non_progressor_longitudnal_only_Sample29	M	Control
GSM2886162	Leicester_non_progressor_longitudnal_only_Sample30	M	Control
GSM2886163	Leicester_non_progressor_longitudnal_only_Sample31	M	Control
GSM2886164	Leicester_non_progressor_longitudnal_only_Sample32	M	Control
GSM2886165	Leicester_non_progressor_longitudnal_only_Sample33	M	Control
GSM2886166	Leicester_non_progressor_longitudnal_only_Sample34	M	Control
GSM2886167	Leicester_non_progressor_longitudnal_only_Sample35	F	LTBI
GSM2886168	Leicester_non_progressor_longitudnal_only_Sample36	F	LTBI
GSM2886169	Leicester_non_progressor_longitudnal_only_Sample37	F	LTBI
GSM2886170	Leicester_non_progressor_longitudnal_only_Sample38	F	LTBI
GSM2886171	Leicester_non_progressor_longitudnal_only_Sample39	F	LTBI
GSM2886172	Leicester_non_progressor_longitudnal_only_Sample40	F	LTBI
GSM2886173	Leicester_non_progressor_longitudnal_only_Sample41	M	Control
GSM2886174	Leicester_non_progressor_longitudnal_only_Sample42	M	Control
GSM2886175	Leicester_non_progressor_longitudnal_only_Sample43	M	Control
GSM2886176	Leicester_non_progressor_longitudnal_only_Sample44	M	Control
GSM2886177	Leicester_non_progressor_longitudnal_only_Sample45	M	Control
GSM2886178	Leicester_non_progressor_longitudnal_only_Sample47	F	LTBI
GSM2886179	Leicester_non_progressor_longitudnal_only_Sample48	F	LTBI
GSM2886180	Leicester_non_progressor_longitudnal_only_Sample49	F	LTBI
GSM2886181	Leicester_non_progressor_longitudnal_only_Sample50	F	LTBI
GSM2886182	Leicester_non_progressor_longitudnal_only_Sample51	M	Control
GSM2886183	Leicester_non_progressor_longitudnal_only_Sample52	M	Control
GSM2886184	Leicester_non_progressor_longitudnal_only_Sample53	M	Control
GSM2886185	Leicester_non_progressor_longitudnal_only_Sample54	M	Control
GSM2886186	Leicester_non_progressor_longitudnal_only_Sample55	M	Control
GSM2886187	Leicester_non_progressor_longitudnal_only_Sample56	M	Control
GSM2886188	Leicester_non_progressor_longitudnal_only_Sample57	M	LTBI
GSM2886189	Leicester_non_progressor_longitudnal_only_Sample58	M	LTBI
GSM2886190	Leicester_non_progressor_longitudnal_only_Sample61	M	Control
GSM2886191	Leicester_non_progressor_longitudnal_only_Sample62	M	Control
GSM2886192	Leicester_non_progressor_longitudnal_only_Sample63	M	Control
GSM2886193	Leicester_non_progressor_longitudnal_only_Sample64	M	Control
GSM2886194	Leicester_non_progressor_longitudnal_only_Sample65	M	Control
GSM2886195	Leicester_non_progressor_longitudnal_only_Sample66	F	Control
GSM2886196	Leicester_non_progressor_longitudnal_only_Sample67	F	Control
GSM2886197	Leicester_non_progressor_longitudnal_only_Sample68	F	Control
GSM2886198	Leicester_non_progressor_longitudnal_only_Sample69	F	Control
GSM2886199	Leicester_non_progressor_longitudnal_only_Sample70	M	Control
GSM2886200	Leicester_non_progressor_longitudnal_only_Sample71	M	Control

GSM2886201	Leicester_non_progressor_longitudnal_only_Sample72	M	Control
GSM2886202	Leicester_non_progressor_longitudnal_only_Sample73	M	Control
GSM2886203	Leicester_non_progressor_longitudnal_only_Sample77	M	LTBI
GSM2886204	Leicester_non_progressor_longitudnal_only_Sample78	M	LTBI
GSM2886205	Leicester_non_progressor_longitudnal_only_Sample79	M	LTBI
GSM2886206	Leicester_non_progressor_longitudnal_only_Sample80	M	LTBI
GSM2886207	Leicester_non_progressor_longitudnal_only_Sample81	M	LTBI
GSM2886208	Leicester_non_progressor_longitudnal_only_Sample83	M	Control
GSM2886209	Leicester_non_progressor_longitudnal_only_Sample84	M	Control
GSM2886210	Leicester_non_progressor_longitudnal_only_Sample85	M	Control
GSM2886211	Leicester_non_progressor_longitudnal_only_Sample86	M	Control
GSM2886212	Leicester_non_progressor_longitudnal_only_Sample87	M	Control
GSM2886213	Leicester_non_progressor_longitudnal_only_Sample88	M	Control
GSM2886214	Leicester_non_progressor_longitudnal_only_Sample92	F	LTBI
GSM2886215	Leicester_non_progressor_longitudnal_only_Sample93	F	LTBI
GSM2886216	Leicester_non_progressor_longitudnal_only_Sample94	F	LTBI
GSM2886217	Leicester_non_progressor_longitudnal_only_Sample95	F	LTBI
GSM2886218	Leicester_non_progressor_longitudnal_only_Sample96	F	LTBI
GSM2886219	Leicester_non_progressor_longitudnal_only_Sample97	M	Control
GSM2886220	Leicester_non_progressor_longitudnal_only_Sample98	M	Control
GSM2886221	Leicester_non_progressor_longitudnal_only_Sample99	M	Control
GSM2886222	Leicester_non_progressor_longitudnal_only_Sample100	M	Control
GSM2886223	Leicester_non_progressor_longitudnal_only_Sample101	F	Control
GSM2886224	Leicester_non_progressor_longitudnal_only_Sample102	F	Control
GSM2886225	Leicester_non_progressor_longitudnal_only_Sample103	F	Control
GSM2886226	Leicester_non_progressor_longitudnal_only_Sample104	F	Control
GSM2886227	Leicester_non_progressor_longitudnal_only_Sample107	M	LTBI
GSM2886228	Leicester_non_progressor_longitudnal_only_Sample108	M	LTBI
GSM2886229	Leicester_non_progressor_longitudnal_only_Sample109	M	LTBI
GSM2886230	Leicester_non_progressor_longitudnal_only_Sample110	M	LTBI
GSM2886231	Leicester_non_progressor_longitudnal_only_Sample111	M	LTBI
GSM2886232	Leicester_non_progressor_longitudnal_only_Sample112	M	LTBI
GSM2886233	Leicester_non_progressor_longitudnal_only_Sample113	M	LTBI
GSM2886234	Leicester_non_progressor_longitudnal_only_Sample114	M	LTBI
GSM2886235	Leicester_non_progressor_longitudnal_only_Sample115	F	LTBI
GSM2886236	Leicester_non_progressor_longitudnal_only_Sample116	F	LTBI
GSM2886237	Leicester_non_progressor_longitudnal_only_Sample117	F	LTBI
GSM2886238	Leicester_non_progressor_longitudnal_only_Sample118	F	LTBI
GSM2886239	Leicester_non_progressor_longitudnal_only_Sample120	F	LTBI
GSM2886240	Leicester_non_progressor_longitudnal_only_Sample121	F	LTBI
GSM2886241	Leicester_non_progressor_longitudnal_only_Sample122	F	LTBI
GSM2886242	Leicester_non_progressor_longitudnal_only_Sample123	F	LTBI
GSM2886243	Leicester_non_progressor_longitudnal_only_Sample125	M	LTBI
GSM2886244	Leicester_non_progressor_longitudnal_only_Sample126	M	LTBI
GSM2886245	Leicester_non_progressor_longitudnal_only_Sample127	M	LTBI
GSM2886246	Leicester_non_progressor_longitudnal_only_Sample129	M	LTBI

GSM2886247	Leicester_non_progressor_longitudnal_only_Sample130	M	LTBI
GSM2886248	Leicester_non_progressor_longitudnal_only_Sample131	M	LTBI
GSM2886249	Leicester_non_progressor_longitudnal_only_Sample132	M	LTBI
GSM2886250	Leicester_non_progressor_longitudnal_only_Sample133	M	LTBI
GSM2886251	Leicester_non_progressor_longitudnal_only_Sample134	M	LTBI
GSM2886252	Leicester_non_progressor_longitudnal_only_Sample135	M	LTBI
GSM2886253	Leicester_non_progressor_longitudnal_only_Sample136	M	LTBI
GSM2886254	Leicester_non_progressor_longitudnal_only_Sample137	M	LTBI
GSM2886255	Leicester_non_progressor_longitudnal_only_Sample140	M	Control
GSM2886256	Leicester_non_progressor_longitudnal_only_Sample141	M	Control
GSM2886257	Leicester_non_progressor_longitudnal_only_Sample142	M	Control
GSM2886258	Leicester_non_progressor_longitudnal_only_Sample143	M	Control
GSM2886259	Leicester_non_progressor_longitudnal_only_Sample144	M	Control
GSM2886260	Leicester_non_progressor_longitudnal_only_Sample145	M	Control
GSM2886261	Leicester_non_progressor_longitudnal_only_Sample146	M	Control
GSM2886262	Leicester_non_progressor_longitudnal_only_Sample150	F	LTBI
GSM2886263	Leicester_non_progressor_longitudnal_only_Sample151	F	LTBI
GSM2886264	Leicester_non_progressor_longitudnal_only_Sample152	F	LTBI
GSM2886265	Leicester_non_progressor_longitudnal_only_Sample153	F	LTBI
GSM2886266	Leicester_non_progressor_longitudnal_only_Sample154	M	LTBI
GSM2886267	Leicester_non_progressor_longitudnal_only_Sample155	M	LTBI
GSM2886268	Leicester_non_progressor_longitudnal_only_Sample156	M	LTBI
GSM2886269	Leicester_non_progressor_longitudnal_only_Sample157	M	LTBI
GSM2886270	Leicester_non_progressor_longitudnal_only_Sample158	M	LTBI
GSM2886271	Leicester_non_progressor_longitudnal_only_Sample159	M	LTBI
GSM2886272	Leicester_non_progressor_longitudnal_only_Sample160	M	LTBI
GSM2886273	Leicester_non_progressor_longitudnal_only_Sample161	M	LTBI

(B)

3.2 Creation of Notebook for the Dataset

With the help of BioJupies tool we have processed the raw RNA-seq dataset into an usable and interactive report. After generating the notebook, export the differentially expressed gene table. The table contained more than 10000 genes which was further filtered on the basis of log FC and adj. P-value to get the upregulated and downregulated genes. Notably, parameters for filtering of log FC value was taken as greater than or equal to '2'or less than or equal to '-2' and for adj. P- value maximum threshold was set at '0.05' [26].

3.3 Exploring Plant-Based Phytochemicals.

The next objective was to find the plant phytochemicals whose genes are matchingwith the dataset. IMPPAT database has enabled the approach of searching the phytochemicals from the medicinal plant [27]. From literature it was predicted that *Allium cepa* (onion) plant has some antimycobacterial properties so in our study also we had firstly referred to plant *Allium cepa* (onion) and some other like *Aloe vera* , *Allium sativum, Azadirachta indica* (neem), *Acalypha indica, Adhatoda vasicaNees*.[28] After getting the phytochemicals and their canonical smiles, structure could be predicted.

3.4 Creation of Notebook for the Dataset

With the help of BioJupies tool we have processed the raw RNA-seq dataset into an usable and interactive report. After generating the notebook, export the differentially expressed gene table. The table contained more than 10000 genes which was further filtered on the basis of log FC and adj. P-value to get the upregulated and downregulated genes. Notably, parameters for filtering of log FC value was taken as greater than or equal to '2'or less than or equal to '-2' and for adj. P- value maximum threshold was set at '0.05' [26].

3.5 Exploring Plant-Based Phytochemicals.

The next objective was to find the plant phytochemicals whose genes are matchingwith the dataset. IMPPAT database has enabled the approach of searching the phytochemicals from the medicinal plant [27]. From literature it was predicted that *Allium cepa* (onion) plant has some antimycobacterial properties so in our study also we had firstly referred to

plant *Allium cepa* (onion) and some other like *Aloe vera* , *Allium sativum*, *Azadirachta indica* (neem), *Acalypha indica, Adhatoda vasicaNees*.[28] After getting the phytochemicals and their canonical smiles, structure could be predicted.

3.6 Target Finding

To predict the most closely related genes/ proteins and targets of small molecules a web tool, Swiss target prediction has been explored[29]. Out of all the plants searched *Allium cepa* (onion) and *Acalypha indica* showed the highest number of target matches. *Allium cepa* had total of 302 matched genes.

3.7 Construction of Protein-Protein Interaction Network Using Matched DEGs

The DEGs that were found matching with the plant phytochemical genes were obtained and given as input of protein protein interaction (PPI). The STRING database was used to construct the protein protein interaction network [30]. The confidence score was set as 0.4 and the network was visualized within STRING database itself.

3.8 Gene Enrichment and Pathway Analysis

The gene enrichment analysis was performed with the help of Enrichr [31]. The enrichment analysis was performed on the phytochemical derived DEGS matching with the selected dataset. Separate enrichment analysis was performed for biological process, cellular component and molecular function. The pathway enrichment analysis was performed with wikipathways.

CHAPTER 4

RESULTS

4.1 Identification of Differentially Expressed Genes in Tuberculosis:

After selection of a suitable RNA-seq dataset from GEO and customizing it as per requirements, we generated a control vs perturbation analysis notebook from BioJupies. The notebook contained several sections like Load Dataset, Clustergrammer, Library Size Analysis, Volcano Plot, Differential Expression Table, Enrichr Links. Our requirement was Differential Expression Table, so we exported it and filtered. After background correction, normalization and filtering with p-value <=0.05 and logFCgreater than or equal to ‘2’ or less than or equal to ‘-2’ the expression data’s, we found 16 upregulated genes and 363 downregulated genes. A basic information of the obtainedresults from the dataset are consolidated into the Table 3. The volcano plot of the datasethas been shown in the Figure 2A where the red marks depict the upregulated genes and the blue marks for down regulated genes., the heatmap representing the gene expressionfor each sample is depicted in 2B. The rows of the heatmap are the genes and every column is the individual samples. The cells represent the expression values after normalization.

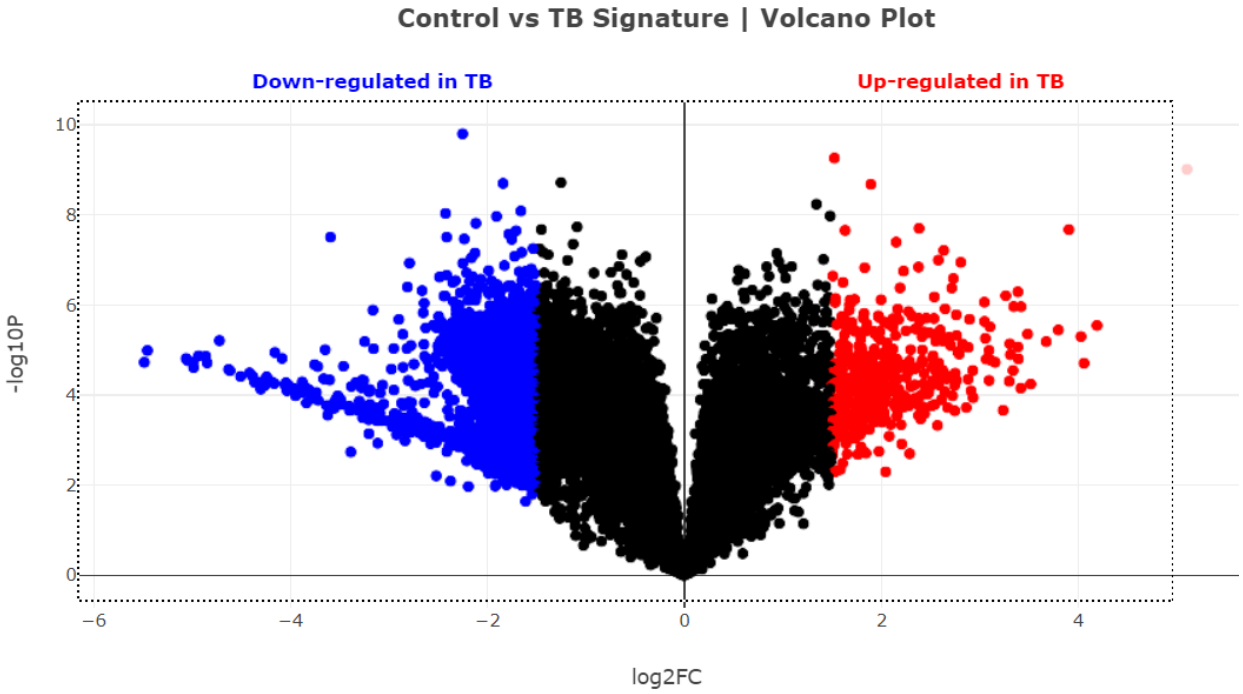


Figure2. Volcano Plot of expression profile of the dataset. (GSE107993)

4.2 Phytochemical Counts

After exploring the IMPPAT database for finding the plant phytochemicals associated genes and then taking their canonical SMILES (Simplified Molecular Input Line Entry System) which is basically a chemical symbolisation of the chemical structures which are computer readable. Those were taken as input to Swiss Target Prediction for matching with the expression gene dataset. Further, we observed that:

- *Acalypha indica* plant phytochemicals had 51 matched genes
- *Alleum cepa*(onion) had 302 genes matching

As the highest number of matches were from onion plant, we chose it for further studies. Figure 3 shows the related phytochemicals that are derived from the onion plant. The network has been created from cytoscape version 4.1

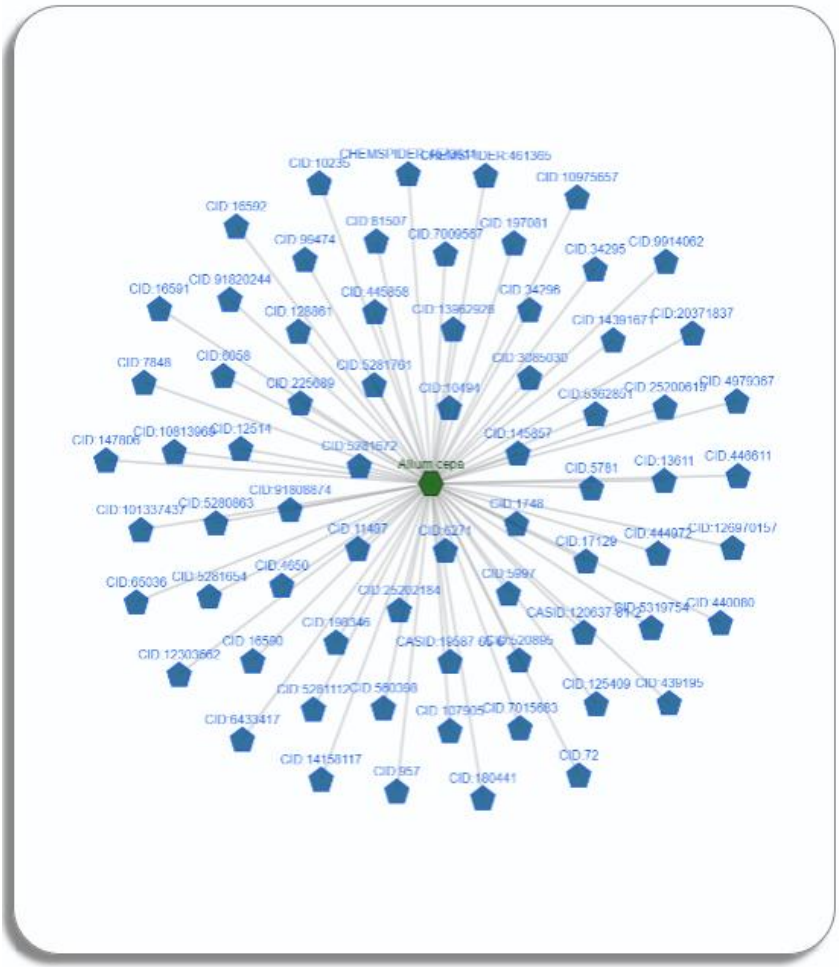


Figure 3: Cytoscape network for medicinal plant (onion) and its associated phytochemicals.

The table 4 below represents the genes matching between the plant phytochemicals and the downloaded dataset from GEO.

S.No.	Phytochemical	Matching Genes
1.	Phytosterols	PTGER2 SMO PTGDR
2.	sinapaldehyde glucoside	ADORA3

		LGALS9 CDC25B TYMP SRD5A1
3.	beta.-Amyrin	CDC25B PTAFR PTGER2 C5AR1
4.	(-)-Epicatechin gallate	FUT4 CA4 ADORA3 ALPL MT-ND4
5.	(1α,2α,3α,4α,5β)-2,3-dimethyl-5,6-dithiabicyclo[2.1.1]hexane 5-oxide	MAP2 PARP10 MPO CDC25B ALOX5 CTSS CTSB ALOX15 ADORA3 SRD5A1 TYMP
6.	(2E)-2-Methylbut-2-en-1-ol	CA4 MPO ALPL CASP1 CYP2D6
7.	1-Caffeoyl-beta-D-glucose	ADORA3 ELANE CA4 CDC25B MGAM TYMP
8.	1-O-(4-Coumaroyl)-beta-D-glucose	CA4 ADORA3 TYMP ELANE CDC25B MGAM
9.	1-O-feruloyl-beta-D-glucose	ADORA3 ELANE ALOX5 CASP1 TYMP
10.	**1-octanol**	CA4 CDC25B

		SPHK1 PTGER2
11.	2-Methyl-2-penten-1-ol	CA4 CASP1 MPO ALPL PTAFR ELANE
12.	2-METHYL-2-PENTENAL	CA4 CASP1 CDC25B PARP10 MPO ELANE ALOX5 SRD5A1 CTSS CTSB ALOX15
13.	2-Methyl-Butyl-2-Methyl-Butyrate	CTSL CTSB ELANE ADORA3 CXCR2 ALOX15 TSPO
14.	2,3-Dimethylthiophene	CTSS CTSL CTSB
15.	2,4-Dihydroxycinnamic acid	ALOX5 CA4 TPMT PTGER2 ELANE ALPL DTYMK

Table 4: List of phytochemicals of fenugreek plant showing matched genes.

4.3 Establishment of Protein-Protein Interaction (PPI) Network

The matching DEGs were taken as input for studying the protein protein interaction network (PPI). The STRING database was used to make the network. The parameter of confidence score was set as '0.4'. The result was visualized in the STRINGdatabase itself. The PPI network is shown in Figure 4.

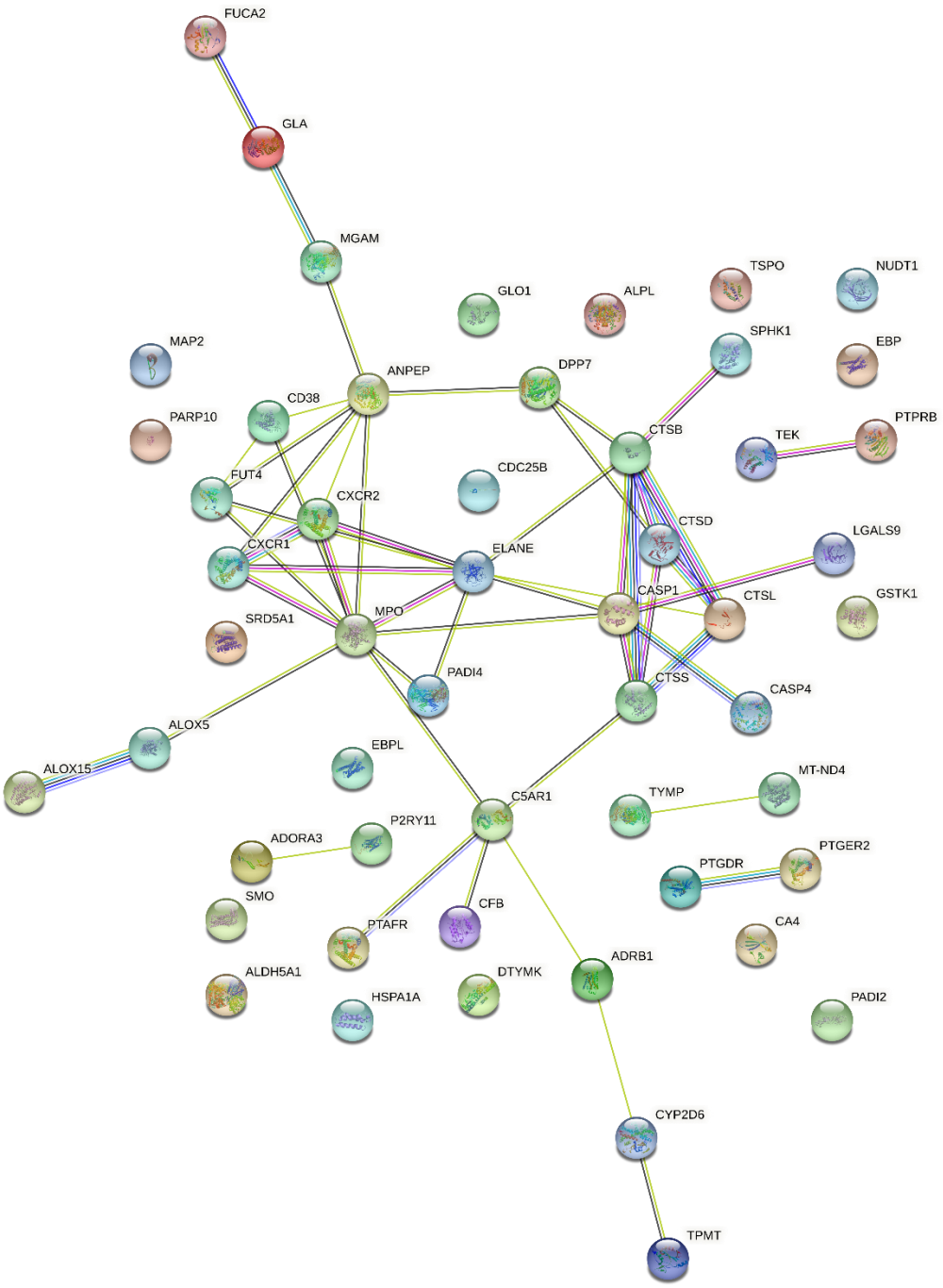


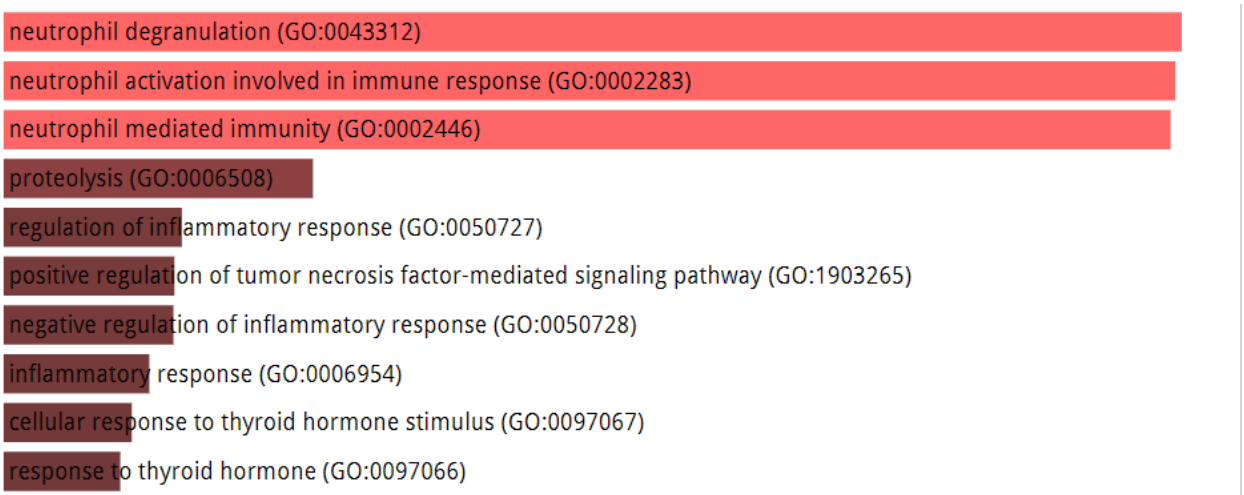
Figure 4: PPI network with the matched

DEGs.

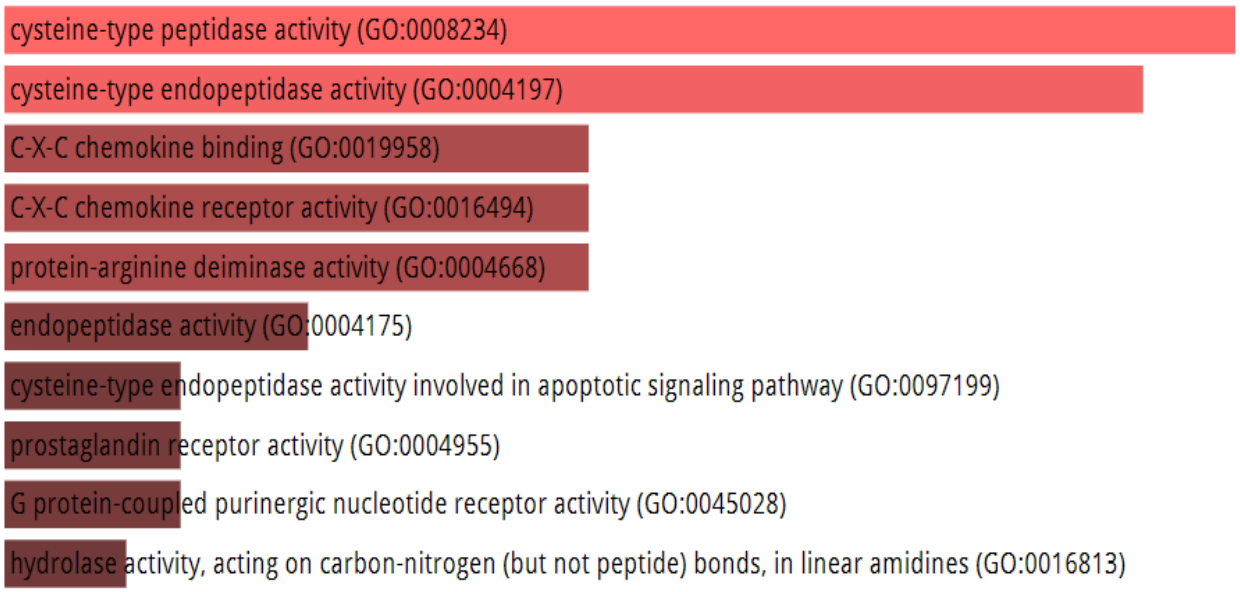
4.4 Gene Enrichment Analysis

The Enrichr tool was used to perform the gene enrichment analysis. The process was performed on the matching genes. The enrichment analysis was performed in biological process, cellular component, molecular function and pathways affected. A Cytoscape plug-in CLUEGO can also be used for the functional enrichment analysis. Forour study we took the help of Enrichr and enrichment analysis was performed with biological process, cellular component, molecular function and wikipathways.

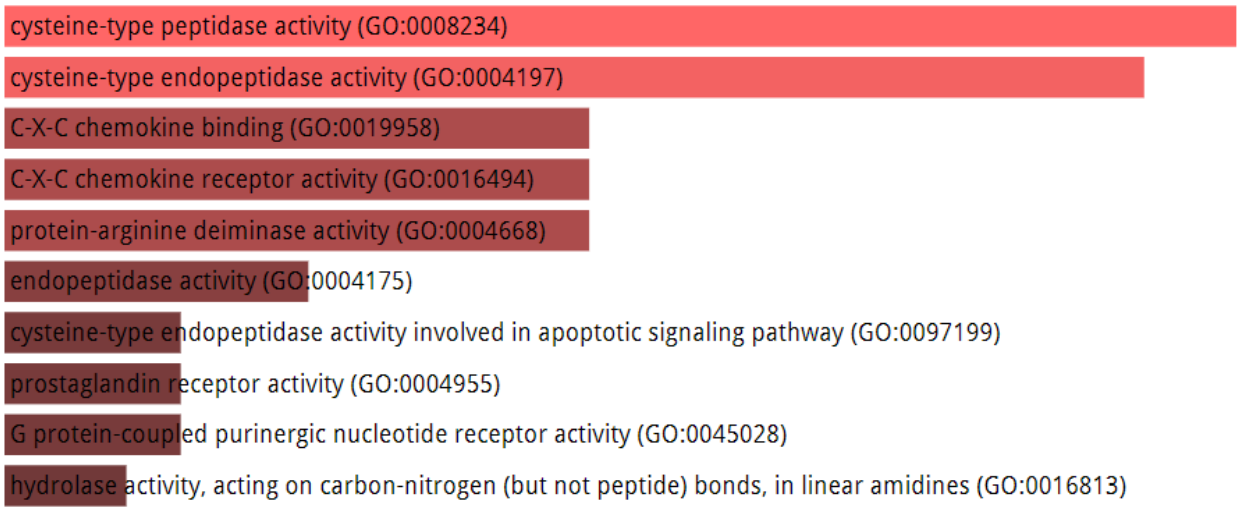
In case of the shortlisted genes, biological processes were over-represented in the neutrophil degranulation, neutrophil activation involoved in immune response, proteolysis, positive regulation of tumor necrosis factor-mediated signaling pathway. The enriched molecular functions were cysteine-type peptidase activity,cysteine- type endopeptidase activity, hydrolase activity. Finally, wikipathways found significantly enriched pathways like Metabolism of alpha-linolenic acid, Nanomaterial-induced inflammasome activation, Small Ligand GPCRs, Benzene metabolism, Melatonin metabolism and effects, PCRs, Biosynthesis with Skeletal Dysplasias, Class A Rhodopsin, Eicosanoid Synthesis. The enrichment result with p value <=0.05 are shown in Figure 5A to 5D.


neutrophil degranulation (GO:0043312)
neutrophil activation involved in immune response (GO:0002283)
neutrophil mediated immunity (GO:0002446)
proteolysis (GO:0006508)
regulation of inflammatory response (GO:0050727)
positive regulation of tumor necrosis factor-mediated signaling pathway (GO:1903265)
negative regulation of inflammatory response (GO:0050728)
inflammatory response (GO:0006954)
cellular response to thyroid hormone stimulus (GO:0097067)
response to thyroid hormone (GO:0097066)

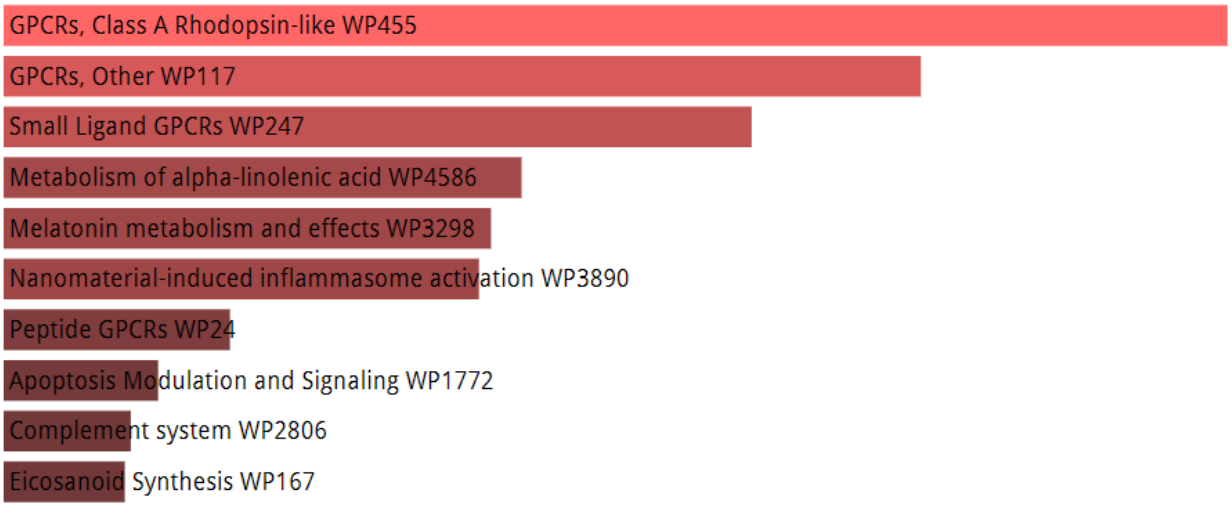
(A)


cysteine-type peptidase activity (GO:0008234)
cysteine-type endopeptidase activity (GO:0004197)
C-X-C chemokine binding (GO:0019958)
C-X-C chemokine receptor activity (GO:0016494)
protein-arginine deiminase activity (GO:0004668)
endopeptidase activity (GO:0004175)
cysteine-type endopeptidase activity involved in apoptotic signaling pathway (GO:0097199)
prostaglandin receptor activity (GO:0004955)
G protein-coupled purinergic nucleotide receptor activity (GO:0045028)
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines (GO:0016813)

(B)



(C)



(D)

Figure 5: Top Gene Ontology (GO) enrichment results along with associated genes in biological process, molecular function, cellular component and the wiki pathway of the phytochemical associated matched genes. **(A-D)**

CHAPTER 5

DISCUSSION

Tuberculosis is considered as one of the critical illness world-wide as ithas a deep negative effect on the quality of life, economic status and mental health of theaffected people. Constant regulation related to physical activity, intake of proper diet, is required. In present, proper diagnosis and treatment of tuberculosis is still not satisfactory, and the risk percentage of people getting affected is still rising. Studies andintensive research on pathogenesis and finding of suitable single nucleotide polymorphism (SNPs) is urgently required as SNPs mainly are considered as good markers as they are directly related to the genes. Some research has reported that SNPs that are situated in the non-coding region are considered more susceptible for diseases. Despite of presently available numerous anti-mycobacterial medicines, their certain side effects and shortcomings related to durability limit their implications. Understanding the reasonfor dysfunction of the β-cells of islet of Langerhans at a molecular level can provide bettertreatment and identification of the pinpointing markers. Some studies found out the marker genes from one of the GEO data-set by undergoingthe steps of exploring DEGs and then network analysis, finally by GO and KEGG enrichment, PPI networking and then functional analysis. In the GEO database there are mainly two types of method used: Microarray and High-throughput sequencing (RNA- seq). In our work preference has been given to high-throughput sequencing method dataas it provides more perfect results in both qualitative and quantitative when compared tomicroarray. After selection of the dataset screening was done by setting some standard parameters to the adj p-value and Log FC columns as they provide the final expression results for controlled and affected samples. In our study we tried to find the related plantwhose phytochemical can have a strong effect on

differentially expression of the genes thus can solve the present challenges. From research it has been found that nutraceuticalsand phytomedicines have comparatively lower side effects thus can be a good alternativefor the drug related complications. Traditionally, using of plant derived medicines, ointments was considered as the best method globally and these approaches has also takenup the credit for its potential as anti-mycobacterial medicine. Hence, the study done on phytochemicals can be considered as a promising way for developing drugs with less complications leading to tuberculosis management.

By gene expression analysis we were able to find the medicinal plants having aneffect for diabetes also upon further target finding could get the phytochemical associatedgenes. Matching them with the GEO selected dataset we were able to get 22 genes from the phytochemicals. From literature, we found that some plants like *Allium cepa, Allium sativum, Aloe vera,* Also, its reported that onion among the top medicinal plant in terms of safety and efficacy. Although research showsthat numerous herbs have the potential of anti-mycobacterial activity. Taking this into account further searched for the plant phytochemicals and their associated genes. Taking this into account further searched for the plant phytochemicals and their associated genes. It was found that onion had the maximum number 291 matched genes. A PPI networking was done to find the inner relation between the genes, keeping the confidence score '0.4' there wasa total of 12 interaction found. TNF and SRC were among the most interacted down- regulated genes and TGM2 was the one gene that was getting upregulated. Then the DEGsisolated were taken for gene enrichment analysis. In biological process, positive regulation of cellular response to organic compound, chemokine binding, chemokine receptor activity, protein-arginine deiminase activity. Integral component of plasma membrane, pseudopodium, chitosome, granule membrane were enriched in cellular component and in molecular function carbonate dehydratase activity, channel activity, ATPase binding, Hydro-lyase

activity, L-amino acid transporter transmembrane activity were found to be over-represented. Analysis of wikipathways identified Metabolism of alpha-linolenic acid, Nanomaterial-induced inflammasome activation, Small Ligand GPCRs, Benzene metabolism, Melatonin metabolism and effects are the highly enriched pathways associated to those DEGs obtained from the matched phytochemical genes.

The enrichment analysis by Enrichr provided helpful information on the phytochemicals and how the differentially expressed genes (DEGs) get regulated by the pathway enrichment. It can be concluded that TNF, SRC are the most regulated genes and also aryl hydrocarbon receptor (AhR) signaling pathway affects the DEGs the most to enhance the metabolism. Not only this, the biological ontologies of cellular response are affected by the downregulated genes. After numerous research it has been found that raw and several extracts from onion helped in controlling tuberculosis. Similarly, in our study we have also found that genes from various molecules/compounds of fenugreek plant having anti-mycobacterial activity.

Currently, only few plants like Aloe vera, onion, garlic and fenugreek has been explored so, exploring the other medicinal plant having anti-diabetic property is the future plan.

CHAPTER 6

CONCLUSION

Disorder like tuberculosis specific competitive genes and medicinal plant- based phytochemicals were identified by gene expression analysis. Also identified the most regulated genes from a very large number of differentially expressed genes. To the best of our knowledge no such studies have been done before on the DEGs from plant phytochemical and their functional and gene enrichment analysis for tuberculosis. In this work, we proposed a methodology for identifying potential biomarkers and plant phytochemicals. Furthermore, we found the most expressed genes and their associated pathways and biological process that can be helpful for further understanding the mechanism underlying pathogenesis of tuberculosis. In depth, such research has not yet been done so more attention is needed in this field. Phytochemicals are plant derived metabolites possessing antioxidant and free radical scavenging activity can be of great therapeutic importance. Recently, plant-based therapeutics is getting light as several evidences are found for its potential todevelop drugs with lesser side-effects as anti-mycobacterial drugs . In future more resource and experiments are needed for the unexplored plants to find their metabolites and effects in developing drugsbefore going for clinical trials just to ascertain any side effects.

References

1. Maurya RK, Bharti S, Krishnan MY. Triacylglycerols: Fuelling the hibernating mycobacterium tuberculosis. Front Cell Infect Microbiol. 2019;9(JAN):1–8.

2. Cole S, R B, J P, T G, C C, D H, et al. Deciphering thebiologyof Mycobacterium tuberculosis from thecompletegenomesequence. Nature. 1998;393(NOVEMBER):537–44.

3. Williams KJ, Boshoff HI, Krishnan N, Gonzales J, Schnappinger D, Robertson BD. The Mycobacterium tuberculosis β-oxidation genes echA5 and fadB3 are dispensable for growth in vitro and in vivo. Tuberculosis. 2011;91(6):549–55.

4. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Prim. 2016;2.

5. Stewart GR, Robertson BD, Young DB. Tuberculosis: A problem with persistence. Nat Rev Microbiol. 2003;1(2):97–105.

6. Silva Miranda M, Breiman A, Allain S, Deknuydt F, Altare F. The tuberculous granuloma: An unsuccessful host defence mechanism providing a safety shelter for the bacteria? Clin Dev Immunol. 2012;2012.

7. Mdluli K, Kaneko T, Upton A. Tuberculosis drug discovery and emerging targets. Ann N Y Acad Sci. 2014;1323(1):56–75.

8. Singh S, Kumar S. Tuberculosis in India: Road to elimination. Vol. 10, International Journal of Preventive Medicine. 2019.

9. Meena LS, Rajni T. Survival mechanisms of pathogenic Mycobacterium tuberculosis H 37Rv. FEBS J. 2010;277(11):2416–27.

10. Davis JM, Ramakrishnan L. The Role of the Granuloma in Expansion and Dissemination of Early Tuberculous Infection. Cell [Internet]. 2009;136(1):37–49. Available from: http://dx.doi.org/10.1016/j.cell.2008.11.014

11. Ndlovu H, Marakalala MJ. Granulomas and inflammation: Host-directed therapies for tuberculosis. Front Immunol. 2016;7(OCT).

12. Pieters J. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. Cell 59 Host Microbe. 2008;3(6):399–407.

13. Daniel J, Maamar H, Deb C, Sirakova TD, Kolattukudy PE. Mycobacterium tuberculosis uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages. PLoS Pathog. 2011;7(6).

14. Jaisinghani N, Dawa S, Singh K, Nandy A, Menon D, Bhandari PD, et al. Necrosis driven triglyceride synthesis primes macrophages for inflammation during Mycobacterium tuberculosis infection. Front Immunol. 2018;9(JUL).

15. Knight M, Braverman J, Asfaha K, Gronert K, Stanley S. Lipid droplet formation in Mycobacterium tuberculosis infected macrophages requires IFN-γ/HIF-1α signaling and supports host defense. PLoS Pathog. 2018;14(1):1–26.

16. Huijing Zhu, Xin Zhu, Yuhong Liu, Fusong Jiang, Miao Chen, Lin Cheng, Xingbo Cheng, "Gene Expression Profiling of Type 2 Diabetes Mellitus by Bioinformatics Analysis", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID9602016, 10 pages, 2020. https://doi.org/10.1155/2020/9602016.

17. Zhang, W., Yu, Y., Hertwig, F. et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome Biol 16, 133 (2015). https://doi.org/10.1186/s13059-015-0694-1.

18. Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: An extended review and a software tool. PLOS ONE 12(12): e0190152. https://doi.org/10.1371/journal.pone.0190152.

19. Grigoriev A. A relationship between gene expression and protein interactions on theproteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. Nucleic Acids Res. 2001 Sep 1;29(17):3513-9. doi: 10.1093/nar/29.17.3513. PMID: 11522820; PMCID: PMC55876.

20. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol Hepatol Bed Bench. 2014 Winter;7(1):17-31. PMID: 25436094; PMCID: PMC4017556.

21. Gonzalez MW, Kann MG. Chapter 4: Protein interactions and disease. PLoS Comput Biol. 2012;8(12):e1002819. doi: 10.1371/journal.pcbi.1002819. Epub 2012 Dec 27.PMID: 23300410; PMCID: PMC3531279.

22. Tiffin N, Andrade-Navarro MA, Perez-Iratxeta C. Linking genes to diseases: it's all in the data. Genome Med. 2009 Aug 7;1(8):77. doi: 10.1186/gm77. PMID: 19678910; PMCID: PMC2768963.

23. Reimand, J., Isserlin, R., Voisin, V. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. NatProtoc 14, 482–517 (2019). https://doi.org/10.1038/s41596-018-0103-9.

24. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107993

25. Singhania A, Verma R, Graham CM, Lee J et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat*

Commun 2018 Jun 19;9(1):2308.PMID: 29921861

26. Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, Iscove N, Jones S, McKinney S, Emerman J, Aparicio S, Marra M, Eaves C. Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell. 2008 Jul 3;3(1):109-18. doi: 10.1016/j.stem.2008.05.018. PMID: 18593563.
27. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samal A. IMPPAT: A curated database of Indian Medicinal Plants,

    Phytochemistry And Therapeutics. Sci Rep. 2018 Mar
28. https://pubmed.ncbi.nlm.nih.gov/17276637
29. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W32-8. doi: 10.1093/nar/gku293. Epub 2014 May 3. PMID: 24792161; PMCID: PMC4086140.
30. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017 Jan 4;45(D1):D362-D368. doi: 10.1093/nar/gkw937. Epub 2016Oct 18. PMID: 27924014; PMCID: PMC5210637.
31. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW,Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1): W90-7. doi:

10.1093/nar/gkw377. Epub 2016 May 3. PMID: 27141961; PMCID: PMC4987924.

32. Salehi-Sahlabadi A, Varkaneh HK, Shahdadian F, Ghaedi E, Nouri M, Singh A, Farhadnejad H, Găman MA, Hekmatdoost A, Mirmiran P. Effects of Phytosterols supplementation on blood glucose, glycosylated hemoglobin (HbA1c) and insulin levelsin humans: a systematic review and meta-analysis of randomized controlled trials. J Diabetes Metab Disord. 2020 Apr 19;19(1):625-632. doi: 10.1007/s40200-020-00526-z.PMID: 32550215; PMCID: PMC7270433.

33. Kazuki Saito, Phytochemical genomicsâ€"a new trend, Current Opinion in Plant Biology, Volume 16, Issue 3, 2013, Pages 373-380, ISSN 1369-5266, https://doi.org/10.1016/.pbi.2013.04.001

34. Zhang YJ, Gan RY, Li S, Zhou Y, Li AN, Xu DP, Li HB. Antioxidant Phytochemicalsfor the Prevention and Treatment of Chronic Diseases. Molecules. 2015 Nov 27;20(12):21138-56. doi: 10.3390/molecules201219753.

35. Alam, Safaet & Sarker, Dr. Md. Moklesur Rahman & Sultana, Taposhi & Chowdhury,Md. Nafees & Rashid, Mohammad & Chaity, Nusrat & Zhao, Chao & Xiao, Jianbo & Hafez, Elsayed & Khan, Shah & Mohamed, Isa. (2022).

turnitin

Similarity Report ID: oid:27535:17841251

PAPER NAME

M.Tech Final Thesis- Shashi Bala Yadav.docx

WORD COUNT	CHARACTER COUNT
5410 Words	**37227 Characters**
PAGE COUNT	FILE SIZE
33 Pages	**1.1MB**
SUBMISSION DATE	REPORT DATE
May 29, 2022 8:36 PM GMT+5:30	**May 29, 2022 8:37 PM GMT+5:30**

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

Excluded from Similarity Report

- Bibliographic material

turnitin

Similarity Report ID: oid:27535:17841251

10% Overall Similarity

Top sources found in the following databases:

- 7% Internet database
- Crossref database
- 5% Submitted Works database
- 6% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	annualreviews.org Internet	1%
2	nature.com Internet	<1%
3	nextias.com Internet	<1%
4	en.zaums.ac.ir Internet	<1%
5	Laredo Community College on 2020-04-17 Submitted works	<1%
6	hdl.handle.net Internet	<1%
7	Sonia Menon, Rodolfo Rossi, Alfred Dusabimana, Natasha Zdraveska, ... Crossref	<1%
8	bmcgenomics.biomedcentral.com Internet	<1%

turnitin

Similarity Report ID: oid:27535:17841251

9	researchsquare.com Internet	<1%
10	assets0.pubget.com Internet	<1%
11	stoptb.org Internet	<1%
12	thieme-connect.com Internet	<1%
13	Saint Leo University on 2011-08-30 Submitted works	<1%
14	repository-tnmgrmu.ac.in Internet	<1%
15	springermedizin.de Internet	<1%
16	Arvind K. Subbaraj, Jan Huege, Karl Fraser, Mingshu Cao, Susanne Ras... Crossref	<1%
17	mafiadoc.com Internet	<1%
18	Anne O'Garra, Paul S. Redford, Finlay W. McNab, Chloe I. Bloom, Robert... Crossref	<1%
19	Zeyu Wu, Xiaowei Ji, Chao Shan, Jie Song, Jin Zhao. "Exploring the pha... Crossref	<1%
20	Nathella P. Kumar, Kadar Moideen, Vaithilingam V. Banurekha, Dina Na... Crossref	<1%

Sources overview

turnitin

Similarity Report ID: oid:27535:17841251

21	brief.land Internet	<1%
22	cibex.nig.ac.jp Internet	<1%
23	maayanlab.cloud Internet	<1%
24	Connie A. Haley. "Treatment of Latent Tuberculosis Infection", America... Crossref	<1%
25	King's College on 2021-06-07 Submitted works	<1%
26	abhipedia.abhimanu.com Internet	<1%
27	slideshare.net Internet	<1%
28	Li Zeng, Kaixue Li, Hong Wei, Jingjing Hu, Lu Jiao, Juan Zhang, Ying Xi... Crossref	<1%
29	Pineda, Silvia, Francisco X. Real, Manolis Kogevinas, Alfredo Carrato, S... Crossref	<1%
30	The University of Buckingham on 2018-09-28 Submitted works	<1%
31	fa.wikipedia.org Internet	<1%
32	Bowling Green State University on 2022-04-08 Submitted works	<1%

Sources overview

turnitin

Similarity Report ID: oid:27535:17841251

33	Kan-Xing Wu, Natalie Jia Ying Yeo, Chun-Yi Ng, Florence Wen Jing Chi... Crossref posted content	<1%
34	thesesups.ups-tlse.fr Internet	<1%
35	Symbiosis International University on 2019-10-21 Submitted works	<1%
36	University of Surrey on 2019-05-07 Submitted works	<1%

Sources overview