IN-SILICO EXPLORING THE EFFECT OF POTENTIAL PHYTOCHEMICALS AS DIABETES THERAPEUTICS

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:

SRIJANI SAMANTA

2k20/IBT/09

Under the supervision of

Prof. YASHA HASIJA

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

MAY 2022

Srijani Samanta

2022

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I, Srijani Samanta, Roll No. 2k20/IBT/09, student of M.Tech in Industrial Biotechnology, hereby declare that the Project Dissertation titled "In-Silico Exploring The Effect Of Potential Phytochemicals As Diabetes Therapeutics" 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 **SRIJANI SAMANTA**

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 "In-Silico Exploring The Effect Of Potential Phytochemicals As Diabetes Therapeutics" which is submitted by Srijani Samanta, Roll No. 2k20/IBT/09, 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.

Place: Delhi
Date:

PROFESSOR YASHA HASIJA
SUPERVISOR

DEPARTMNT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi - 110042

ABSTRACT

Diabetes mellitus, a serious metabolic disease caused by an unbalanced blood glucose level, continues to be a significant global burden, despite the availability of numerous treatments to manage the level. As a result, there is an essential need for an innovative approach to diabetes treatment. In research, Diabetes mellitus has been connected to the generation of free radicals and its antioxidant capacity. Phytochemicals, which are plant-derived metabolites that act as antioxidants and free radical scavengers, 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 diabetes management. We analyzed a RNAseq 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 responses to 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 Rajkumar 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 **SRIJANI SAMANTA**

Date:

CONTENTS

LIST OF FIGURES

LIST OF TABLES

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. Diabetes, Cancer, Stroke, Arthritis are some well-known chronic disorders.[1]

Diabetes is considered as a common multifactor regulated disorder worldwide. Studies revels that the percentage of affected in both men and women will increase with upcoming times. Looking back, it has been found that diabetes has almost doubled the death rate [1] and has become the 7^{th} leading factor for death globally.[2] Diabetes is mainly caused due to non-functionality of β-cells of pancreas secreting insulin or even if the insulin is produced the body cells are not capable of taking up.[3] Various groups of researchers due to chronic nature of diabetes carried out multiple experiments and found that DM is corelated with high formation of free radicals and lowering the antioxidant potential. [4] Due to certain shortcomings related to durability and side effects of presently prescribed anti-diabetic medicines, every year numerous deaths occur worldwide. Majorly diabetes is characterized into three major groups: [5]

- Type I (T1DM) – also named as Insulin Dependent Diabetes Mellitus (IDDM). Around 10% of population with diabetes is found to have type 1 diabetes. It's an autoimmune disorder due to failure of pancreas to secrete insulin rightly as it lacks β-cells. Proper cause for such happening is not yet known but are guessed to be somehow having relation with genes and environmental factors. It is said to be that having a family member affected rises the risk for getting diabetes. Some symptoms like losing weight and energy, recurrent urination tendency when seen in patients in-vitro supply of insulin is provided for the survival of patient. Type 1 is more common in children and young group of population.
- Type II (T2DM) -This condition is reported to be more common in adults and almost 90% of population with diabetes is found to be having type 2. Due to

obesity, there becomes a condition where the body cells cannot property function towards the insulin produced. Patients are asked to have proper diet and do more physical activity however, drugs and insulin is provide for maintenance of glucose levels.

- Gestational (GDM) - This is mainly found in women during their pregnancy and is linked to both mother and infant. Later on, GDM disappears but still remains a risk of developing type 2 diabetes.

Diabetes is considered to have multiple genetic loci that are crucial for determining risk. Finding the transcriptional changes within the whole genome by the help of various molecular biology tools like microarray, high throughput sequencing helped the biomedical sector to grow. Some research tells that the polymorphism in the genes and signalling pathways results in the risk for T2DM. So, understanding the background mechanism for development of diabetes is most crucial before going into medication and therapy.

Glucose being the active regulator for diabetes it moderates the insulin gene expression and secretion. Gene expression is an organic-phenomena by which a gene is regulated within a cell to form RNA and proteins [6]. 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 massive progress in high throughput sequencing gave rise to a new domain of 'phytochemical genomics' [7]. 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.[8] Thus, phytotherapies are gaining attention nowadays for treating diabetes and finding new drug model. Antidiabetic phytochemicals taken from medicinal plants provides identical ways for developing functional foods and antidiabetic drugs.[9] The role of bioactive phytochemicals as antidiabetic drug development has been represented in Figure 1 below.

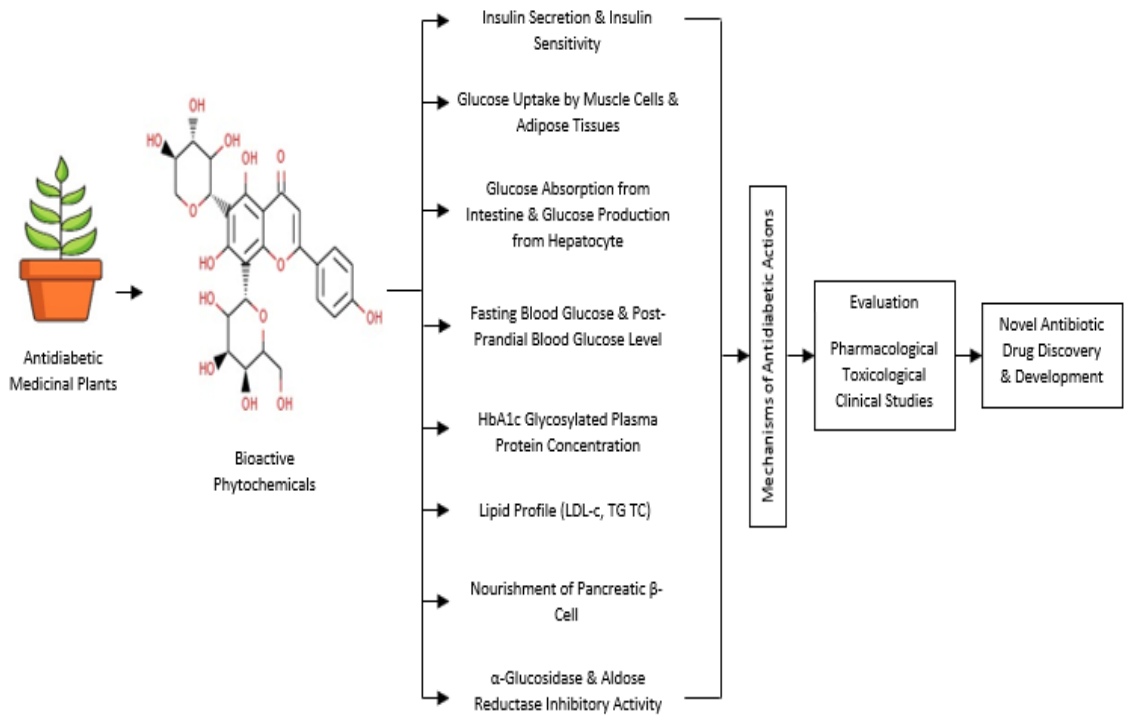


Figure 1: Flowchart on phytochemical-based drug development process.

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.[10] It has been reported that plants like ginger, tea and fenugreek and phytochemicals (i.e. curcumin and resveratrol) with an impact on diabetes have been shown to interact with different protein targets of T2DM. [11]

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 repository for 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

Diabetes Mellitus (DM) is a well-known chronic developing oxidative stress disorder causing destruction of blood vessels which can further lead to stroke and cardiac arrest also results in problems like kidney dysfunction, loss of sight.[12] The primary cause of death in the world is due to diabetes and its complications. The risk for diabetes is increasing in both developed and developed countries. As the name chronic suggests chronic is a long-lasting illness and due to imbalance in blood pressure and blood glucose levels it puts further stress on heart. Unlike other chronic disorders, the primary symptoms of diabetes slow injury healing and high tendency of urination.[13] There is no particular age for onset of diabetes but usually seen that the type 2 starts over the age of 45, thus it is termed as 'adult-onset' diabetes. It has been reported that DM is mainly associated with high formation of free radicals and low antioxidant potential Immune system destructing the cells of pancreas which secrete insulin and also improper diet and overweight are the main cause for diabetes.[14] It has also been proved earlier that certain factor like high inflammation stimulates the tissue injury and vascular damage leading to several complications such as diabetes, diabetic neuropathy, kidney diseases.[15] Glucose, a monosaccharide whose concentration in blood stream is one of the major regulator for increasing the risk for diabetes. The three major ways of circulating it inside body are intestinal absorption, glucogenesis and glycogenesis.[16] Glucose Oxidation is one of the prominent factor for generating free radicals. We all know that glucose is the main source of energy in our body and it's metabolism is mainly regulated by insulin as it is responsible for allowing glucose to enter the cells and inhibition of gluconeogenesis, lipolysis, ketogenesis and proteolysis.[17] Thus, it is considered as the primary therapeutic target for diabetes affected patients. In case of type 1 diabetes patients, pancreas is unable to secrete insulin resulting in accumulation of glucose as in absence of insulin glucose is not taken up into cells whereas, in type 2 patients pancreas releases less insulin than actually needed. So, as a result glucose is not utilised as an energy substituent

leading to serious heath complications.[18] Any person with a BMI>25 and also above 45 years posses greater risk towards diabetes. It is considered that any person with 60-140 milligram of sugar/deciliter of blood is normal/ healthy. If the reading fluctuates above 140mg/dl then it is said to be the condition of hyperglycemia i.e. the insulin doesn't work properly and the body is unable to utilize the glucose at the fullest. Hyperglycemia along with hyperlipidemia activates cytokines (IL-6, TNF-α) and chemokines (CCL1, CXCL1).[19] Also, if the mark is below 60mg/dl then it is too low and is termed as hypoglycemia. Hypoglycemic patients are suggested to carry syringe containing glucagon as this hormone stimulates liver to secrete more sugar.[20] Although both the types have different causes but have something in common which is both type patients inherit a predisposition of the disease from their parents which gradually is expressed due to certain environmental conditions like weather, early diet, virus.[21] Still its sometimes reported that type 2 has a stronger link to be getting inherited than type 1. Diabetes research has an elongated past of assessing genes for disease correlation in case-control and family-related studies.[22] However, only a few chromosomal areas regularly showed major diabetes association across various projects. Some of them are, association between regions of Human Leukocyte Antigen (HLA) on chromosome 11p15 and also the cytotoxic T-lymphocyte in combination with (CTLA4) region on chromosome 2q33, a 4kb stretch at 9p21.3 region. The main genetic disorders which result to diabetes risk are: β-cell non-functioning leads to diabetes in young one's and also changes in mitochondrial DNA.[23] During insulin processing/action defects are caused in conversion to proinsulin due to respective gene and receptor mutation. The genetic testing for diabetes is helpful for analysis caused by point mutation in the genes and also targeted genotyping plays a key role for diagnostic purpose of the disease. As found by the researcher's missense variant of a gene increased the risk of diabetes by almost 5folds. [24] Candidate gene studies were previously used to identify the risk genes. This method usually starts with the selection of candidate genes depending on its functional role and involvement in metabolic pathway and then identifying the related SNPs. Then these SNPs were used to genotype in a random population containing cases and controls followed by final testing to find associations. But maximum times consistent results related to associations were only observed so researchers were not fully satisfied with this approach [25] and hence, moved for GWAS (Genome Wise Association Study) to finalize the genetic element for Diabetes. Via this method, no initial data about core biology or risk alleles is necessary. This method usually scans through the entire Genome in a certain

population for finding the Single Nucleotide Polymorphism (SNPs). This method basically uses chip microarray technologies such as Illumina. By the year 2009 GWAS was successful in identifying 38 SNPs having link with T2D from different population.[26] The TCF7L2 was considered to be the most prominent transcription factor gene associated with diabetes by altering glucose homeostasis and also some other like obesity. Thus, considered as a pleiotropy (multiple effects in one common genetic loci) also making diabetes as a complex genetic disorder. TCF7L2 is also reported as a crucial component of WNT pathway as it helps in proliferating β-cells for insulin secretion and action.[27]

The advancement of technologies and GWAS studies made it easier for the researchers to undergo further studies on pharmacogenomics and development of safety drugs to cure diseases. This also helped them to get a deeper understanding about the patient response for a drug and were able to solve the previously faced challenges. With time, researchers identified more SNPs in the genes like CYP2C9, KCNJ11, PPARG and also their association with insulin signalling like pathways.[28] GWAS studies also allowed in finding how certain variants are responsible for the effect of drugs in individuals thus, helped the clinicians in making proper decision and give efficient dosage for treatment and management of diabetes. Although after all research only 10% variants were explained by GWAS and the other 90% are said to be in the non-coding region whose search is still going on by the help of Next Generation Sequencing method.[29] For diabetes research 'epigenetics' also played a crucial role by providing the idea about molecular links between diabetes and genetics as it plays important part in cellular processes. It has also been reported that epigenetics impacts the protein-protein interactions that are related to insulin secretion pathway for example the genes GLP1 and PAX4.[30] Another part of omics is proteomics that particularly deals with study on mechanisms and find novel markers for early detection and cure. Different research reported about the proteins and their regulation (up-regulated/ down-regulated) when collected from different samples (urine, plasma) from patients. Thus, together genomics and proteomics gives a boost to the medical sector researchers as it helps in giving clarity about the molecular basis of disease and help in improving the diagnosis and early treatment.[31]

Currently, medicinal plants and phytochemicals as name suggests 'phyto' means plants, basically bioactive nutrients derived from plants in the form of fruit, grains,

vegetable, beans are gaining attention of the scientists as they have several health benefits. Preclinical and clinical studies have reported that phytochemicals have the potency of curing various diseases due to its antioxidant and anti-inflammatory roles.[32] Trials on animals have showed its effect of reducing the risk of disease that are mainly linked with oxidative damage. In natural antioxidants are mainly classified as primary antioxidants which act as free radical terminators, secondary helps in retarding chain initiation and tertiary helps in repairing damaged biomolecules.[33] Antioxidants has the ability to remove free radicals present in the human body and gives a protection on diseases such as Cardiovascular-disease (CVD), Cancer and Type 2 Diabetes (T2D). Phytochemicals such as flavonoids, phytoestrogens, phytosterols and fibres are reported to provide more health benefits. Also, at the same time phytochemicals contains some poisonous and toxic chemicals too, thus proper dosage of it is necessary. Phytochemicals derived from herbs containing bioactive compounds have major hypoglycemic effects and is found to play a major role in diabetes linked vascular challenges with less toxicity and decreased side effects.[34] Usage of medicine derived from plants for treating diabetes is already effective in Africa, America and few parts of Asia. As per reports of WHO in 2008 above 50% patients suffering from diabetes depends on herbal medicine prepared from medicinal phytochemicals.[35] Studies have reported that certain phytochemicals like (flavonoids, lignans, monoterpenes, phenylpropanoids) protect the oxidative stress related disorders like diabetes and also prevent the formation of advanced glycated end products (AGEs). Thus, we can say phytochemicals are naturally derived which changes the action of insulin and has the potential of antidiabetic agent. Instead of having numerous biological health benefits against diabetes only a few studies have been done till date. So, it is recommended to explore more about different plants therapeutic effect. Some recent work on phytochemicals for treatment of diabetes shows results like Momordicine I and II activates insulin secretion from beta cells, Trans-tiliroside from plant Potentilla chinesis helps in decreasing glucose and cholestrol concentration and triglyceride levels from diabetic samples, Kaempferol-3-neohesperidoside is an effective antidiabetic compound which shows insulin mimic roles, Bergenin isolated from plant Caesalpinia digyna showed antidiabetic and antioxidant activity also posses a good effect on pancreatic cells, Marrubin from Leonotis leonurus helped in improving insulin levels and gene expression of glucose transporter in INS-1 cells, Alisol F and B compounds of plant Alismatis Rhizoma showed in-vitro glucosidase inhibitory activity, Iridoid glycosides, Ningposide I and II found from plant Scrophularia ningpoensis gives alpha-

glucosidase inhibitory action, Malonyl ginsenosides from plant Panax ginseng gives low blood glucose concentration and improves insulin sensitivity and lipid level in diabetic samples, 6-O-galloyl-5'-hydroxyl mangiferin, methyl gallate are isolated from plant Mangifera indica helps in reducing blood glucose levels in animal diabetes samples, Ginsenoside Re, showed anti-diabetic activities by lowering insulin resistance in PPAR-γ pathways, Chicoric acid that is taken from plant Ocimum gratissimum L. helped in lowering glucose levels in diabetic samples, Asiatic acid shows anti-diabetic role by improving lipid in rats.[36]

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 in genes 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.[37] It is important in understanding the biological differences between a healthy and a diseased person. DGE is associated with methylation of DNA.[37] 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).[38] A different method used is 'exome sequencing' where only particular nucleotide changes in the coding regions that results to a specific phenotype. Recently, research find this technique more useful. edgeR, DESeq2, R package are some well-known software used for differential gene expression analysis.[39]

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.[40] 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 all the proteins used for predicting the network between them some are called 'hub proteins' as they show high number of interactions.[41] 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. [42] In a study it has been found 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.[43] Once the networks are clearly understood, drug designing becomes easy for the researchers as they can easily identify the potential drug targets. 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.[44] 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 [45] All the tissue samples were obtained from skin of both type 2 diabetes patients and non-diabetes patients. Basic information about the dataset is shown in Table 1A and 1B.

Table 1: Brief information on the selected dataset.

Tissue Origin	Geo Accession Number	Sequencing Technique Used	Sample Size	
			Control	Patient
Dermal Endothelial Cells (Human Skin)	GSE92724 [46]	RNA Sequencing OR High Throughout Sequencing	6	4

(A)

Sample Geo Accession	Sample Title	Cell Type	Disease State	Gender
GSM2436514	DM.Ctrl 1	Endothelial Cell	Healthy Control	Female
GSM2436515	DM.Ctrl 1	Endothelial Cell	Diabetic Patient	Male
GSM2436516	DM.Ctrl 2	Endothelial Cell	Diabetic Patient	Female
GSM2436517	DM.Ctrl 3	Endothelial Cell	Diabetic Patient	Female
GSM2436518	DM.Ctrl 4	Endothelial Cell	Diabetic Patient	Male
GSM2436519	DM.Ctrl 2	Endothelial Cell	Healthy Control	Female
GSM2436520	DM.Ctrl 3	Endothelial Cell	Healthy Control	Female
GSM2436521	DM.Ctrl 4	Endothelial Cell	Healthy Control	Female
GSM2436522	DM.Ctrl 5	Endothelial Cell	Healthy Control	Female
GSM2436523	DM.Ctrl 6	Endothelial Cell	Healthy Control	Female

(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 total of 379 genes which was further filtered on the basis of logFC and adj. P-value to get the upregulated and downregulated genes. Notably, parameters for filtering of logFC 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'.[47]

3.3 Exploring Plant-Based Phytochemicals.

The next objective was to find the plant phytochemicals whose genes are matching with the dataset. IMPPAT database has enabled the approach of searching the phytochemicals from the medicinal plant.[48] From literature it was predicted that Aloe vera plant has some anti-diabetic properties so in our study also we had firstly referred to *Aloe vera* plant and some other like *Allium cepa* (onion), *Cinnamomum cassia*, *Azadirachta indica* (neem), *Trigonella foenum-graecum* (Fenugreek).[49] After getting the phytochemicals and their canonical smiles, structure could be predicted.

3.4 Target Finding

To predict the most closely related genes/ proteins and targets of small molecules a web tool, Swiss target prediction has been explored.[50] Out of all the plants searched *Aloe vera* and *Trigonella foenum-graecum* (Fenugreek) showed the highest number of target matches. Fenugreek had total of 22 matched genes.

3.5 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.[51] The confidence score was set as 0.4 and the network was visualized within STRING database itself.

3.6 Gene Enrichment and Pathway Analysis

The gene enrichment analysis was performed with the help of Enrichr.[52] 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

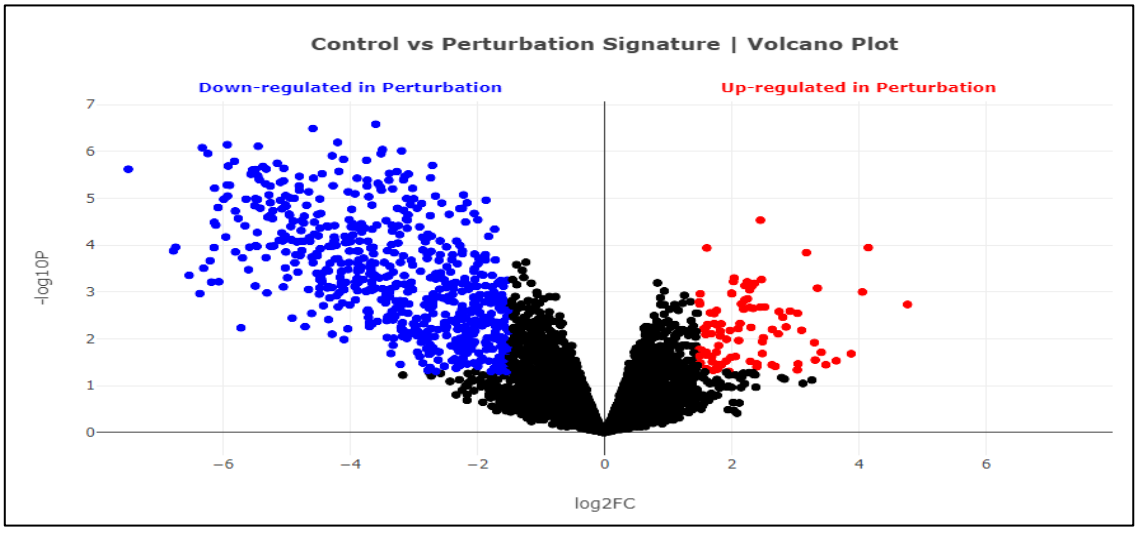
4.1 Identification of Differentially Expressed Genes in Diabetes:

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 Datasset, 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 logFC greater 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 obtained results from the dataset are consolidated into the Table 3. The volcano plot of the dataset has 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 expression for 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.

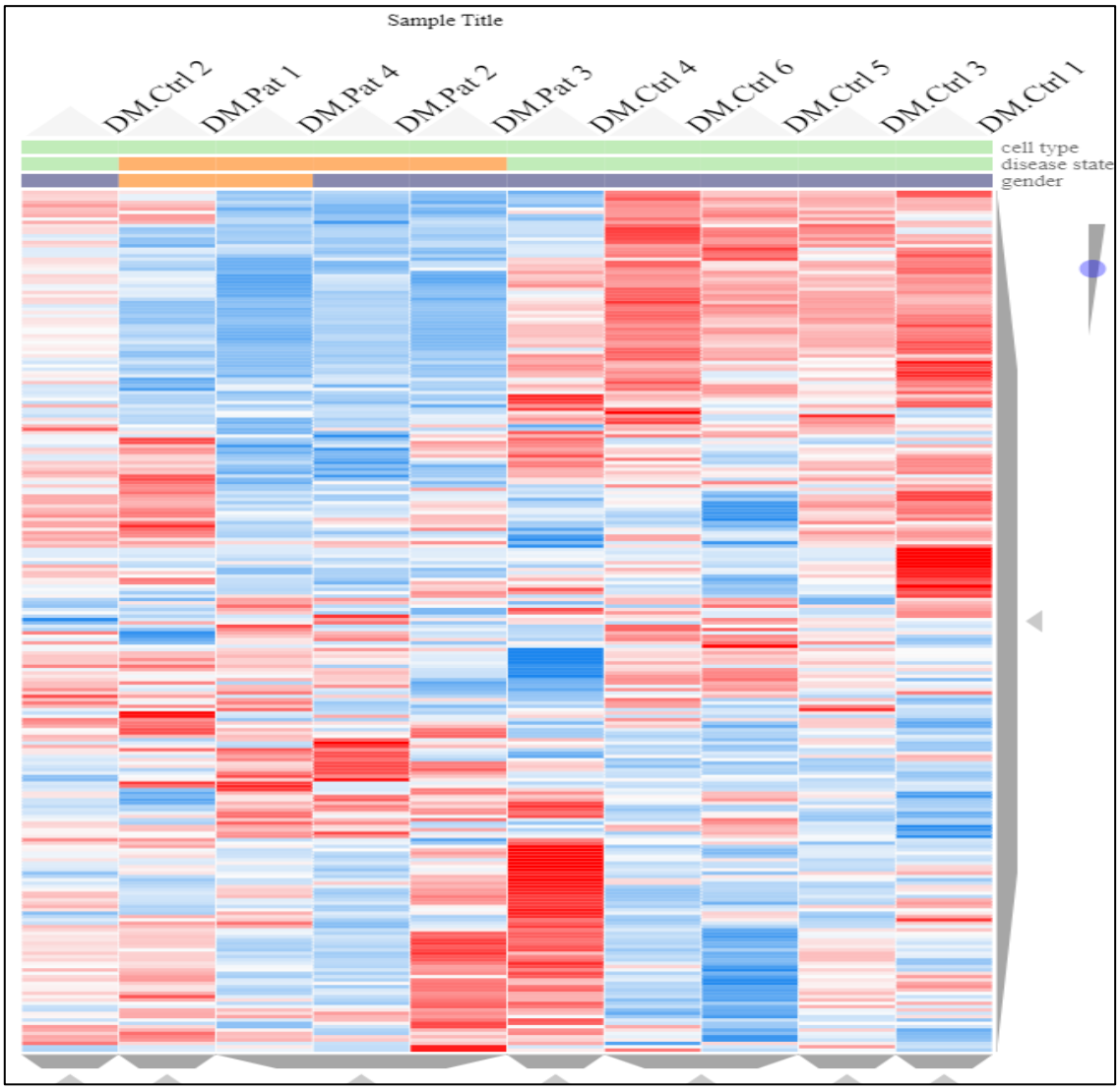
Table 2: Obtained DEGs from the dataset.

DOWN-REGULATED	GRTP1, FAM110C, IRX5, LRP3, TOX, TTC22, KRT77, ADORA2B, MXRA5, RBM11, TCEAL2, PLEKHN1, FOXQ1, CHPF, NCF2, FAM83H, OVOL1, RIC3, SERPINB2, LGALS7, CALML3, IL7, EPHX3, THEM5, RAB27B, TFAP2B, SLURP1, BCL2, CPNE7, ZNF703, GRHL3, SULT1E1, CLEC12B, BNC1, AR, CPA4, BICDL1, CLEC11A, SUSD2, PTPRZ1, ITGB6, PIK3C2G, LYPD3, GAL3ST4, FOXN1, FAM83B, HLA-DQB2, GRIP1, KLK8, ID4, COL6A2, SCEL, PDGFC, SOX15, RRAGD, TOM1L1, UST, TRPA1, FAM134B, SCML2, NIPAL1, HS3ST6, GPR27, SRC, LRP1, LGR4, WNT7B, SFN, CKMT1B, EDIL3, LYPD6B ,ELOVL4, MLANA ,OVOL2,, LMO1, APCDD1, BBOX1, SPON2, NDUFA4L2, PTPRF, ABCA12, FAM83C, LAD1, CELSR2, SPTSSB, ANLN, CYP2C18, PPP2R2C, PCSK2, PAK6, KRT16P6, DSG1, POU3F1, CBLC, GRHL2, GSDMC, CRYAB, SOSTDC1, EPHB6, FAM83A, HCAR2, TPPP3, C1ORF106, S100A14, CA2, EPB41L4B, FRG2HP, ZNF711, EFS, PLA2G4F, TFCP2L1, KRTDAP, ASPN, FERMT1, AXIN2, GATA3, LGALS7B, IGSF9, CSTA, IGSF3, PMEL, PRLR, FZD7, KCNK1, PLEKHG3, DUOXA1, SOX9, CNTN1, ARHGEF10L, SLC24A5, SEPT5, S1PR5, BASP1, BMP7, RNF128, KRT14, CLDN4,

	DLK2, VSNL1, SLC7A5, ALDH3B2, ST14, SLC6A9, DAPP1, FGFBP1, LY6D, NECTIN4, SDC1, MYCL, KRT31, SMOC2, PPP1R14C, BTBD11, DDR1, AQP3, ERBB3, AP1M2, KBTBD12, AADACL2, ANKRD22, RGS1, TRIM29, NOTCH3, CRABP2, SLC39A2, LY6G6C, WNK2, KRT1, SULT2B1, GJB3, NSG1, HMGCS2, SPINT1, KRT5, LAMB4, SERPINB5, MAL2, EXPH5, CDHR1, PLP1, TP53AIP1, PKP3, TNF, HOXC10, CLDN1, FAM57A, KRT80, P2RY1, MSX2, IRF6, TNNT1, DMKN, CHL1, AMPD3, KDF1, GNA15, UGT1A6, DSC1, DEFB1, EPS8L2, DAPL1, CDC42BPG, STK26, CDCA7L, HOOK1, HCAR3, SLC22A15, NIPAL4, DLX3, MYH14, CD44, CLIP4, PKP1, ANKRD35, RYR2, FXYD3, LGI3, CASP14, COBL, SPTBN2, EDARADD, IL1RN, RHOV, TYRP1, CDS1, TYR, MAF, FGFR3, WNT4, GJB6, LCH2, CXCL14, ANXA8, PKIA, RAB25, ESRP1, CAMSAP3, SYT8, VANGL2, IL20, PSAT1, BCL11A, TGFBI, CXADR, ANXA8L1, ANO1, CLCA2, TNNI2, TINCR, RIPK4, IRX3, NOTCH2, POU2F3, RYR1, S100A2, DSC3, EHF, DLX5, TRNP1, CEBPA, PLEK2, GPR87, CKB, OLFM2, MBP, FCER1A, SLC2A1, AC007325.2, CA12, LARGE2, CDH1, NEO1, TNFRSF18, DSE, ESRP2, MAPK13, MPP7, ADGRF4, CLCA4, MFSD2A, TFAP2C, EVPL, PHGDH, NCS1, TMEM30B, CDH3, RAPGEFL1, GATM, CLEC2A, IRX2, IL20RA, WNT16, MOXD1, NUDT11, COL16A1, HPGD, RND3, FAM160A1, MUC15, DCT, ARHGEF4, CERS3, CCL27, SLC22A17, PARD6G, ANTXR1, KRT10, LRP4, ANKRD33B, IRX4, CHP2, KLK11, CDR1, IL18, KREMEN2, SH3RF2, CAPNS2, TNFRSF19, CKMT1A, WFDC5, HAS1, VAV3, TFAP2A, ODF3L1, IL20RB, LSP1, SPINT2, CDCA7, FGFR2, GPNMB, KLF5, BNIPL, GJB5, GDPD2, FAM46B, SDC4, WNT3, SLC1A3, CDCP1, SEMA3C, SLITRK6, BCL11B, SERPINB7, KRT15, FZD10, EPHB3, NRG1, TNS4, NMU, FMO1, HR, COL7A1, PRSS8, DSG3, CD1A, KLC3, PTPN13, CD207, TP63, FAT2, ZNF750, F2RL1, APOE
UP-REGULATED	NDRG4, CPE, FN1, KCNJ12, FGF18, ISLR2, GJA4, OLFML3, CYP1B1, LYPD6, ZSCAN4, KCNK15, RPS16P9, ACKR3, G0S2, TGM2



(A) Volcano Plot of expression profile of the dataset. (GSE92724)



(B) Heatmap showing gene expression for every sample.

Figure 2: Processing of dataset obtained from GEO. **(A)** Volcano plot showing upregulated and downregulated genes. **(B)** Heatmap showing expression level of each sample.

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:

- *Aloe vera* plant phytochemicals had 17 matched genes
- *Alleum cepa* had 8 genes matching
- *Trigonella foenum-graecum* (Fenugreek) phytochemicals had 22 matched genes.
- Neem plant phytochemicals had 12 matched genes.

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

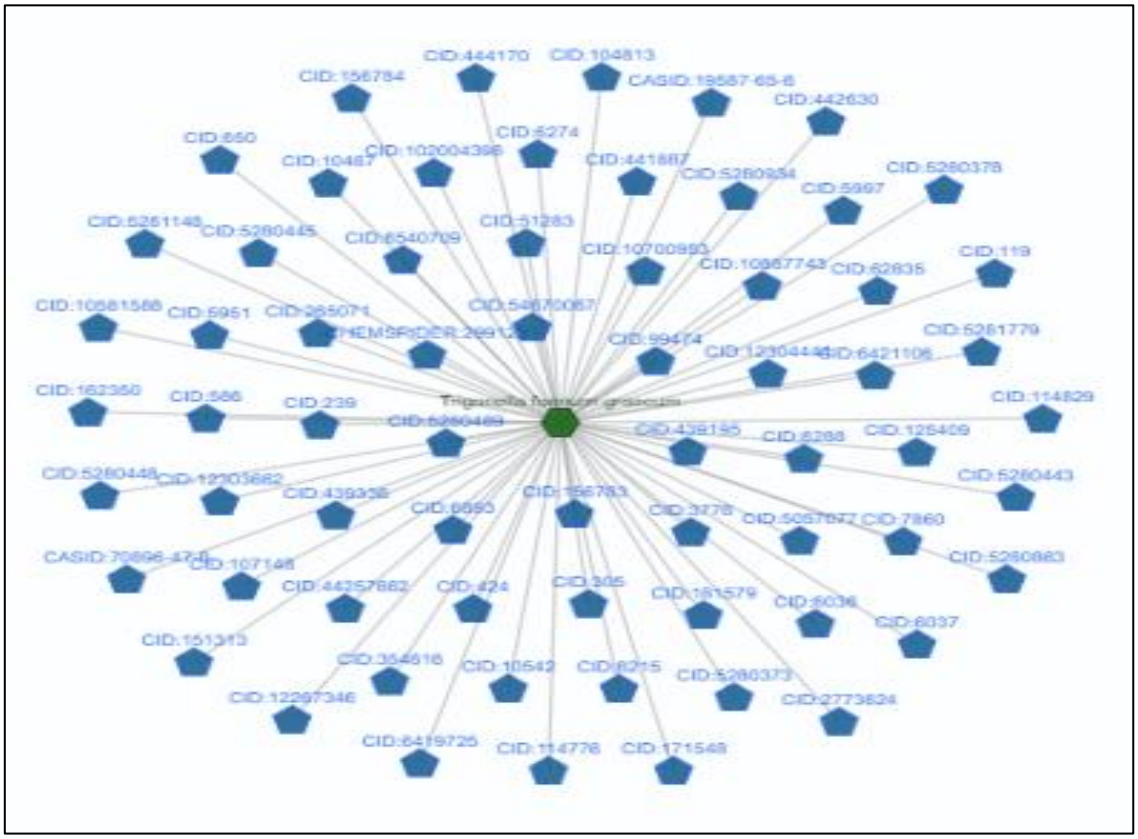


Figure 3: Cytoscape network for medicinal plant (fenugreek) and its associated phytochemicals.

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

Table 3: List of phytochemicals of fenugreek plant showing matched genes.

SL.NO	PHYTOCHEMICAL NAME	GENE MATCHING
1	Phytosterols	AR, PTPRF, TNF
2	Spirostan-3-ol, (3beta,5beta)-	ADORA2B, AR
3	Vicenin 1	CA12
4	(25R)-Spirosta-3,5-dien	TGM, SLC6A9, TGM2
5	2,3-butanedione	TRPA1, CA12,
6	2,3-Dimethylaniline	CA2
7	2,4,4'-Trihydroxychalcone	TYR, AR, BCL2
8	3-Amino-4,5-dimethyl-2(5H)-furanone	CA2, CA12, TGM2, ADORA2B, TRPA1, HCAR2
9	3,4,7-trimethylcoumarin	CYP1B1, CA12, TGM2, CA2
10	4-aminobutyric acid	BBOX1, CA2, CA12, AR
11	4-Hydroxyisoleucine	BBOX1, AR, CA2, CA12, SLC1A3
12	45-Dimethyl-3-Hydroxy-25h-Furanones	HCAR2, CA2,
13	Apigenin	CYP1B1, CA2, CA12, TYR, AR, SRC
14	Gitogenin	AMPD3, SRC, TNF, CA2, ADORA2B
15	l-ascorbic acid	AR, F2RL1, CA2
16	D-Raffinose	P2RY1, CA2, CA12, TYR
17	Fenugreekine	CA2, P2RY1, CA12, LGALS7, TYR, BCL2
18	Folic acid	BCL2, PTPRF, ADORA2B, TNF CA2, AMPD3, PIK3C2G
19	Choline	HPGD

Also, after matching it is seen that most of the matched genes are getting downregulated and only one from the list i.e. TGM2 was getting upregulated.

4.3 Establishment of Protein-Protein Interaction (PPI) Network

The matching 22 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 STRING database itself. The PPI network is shown in Figure 4.

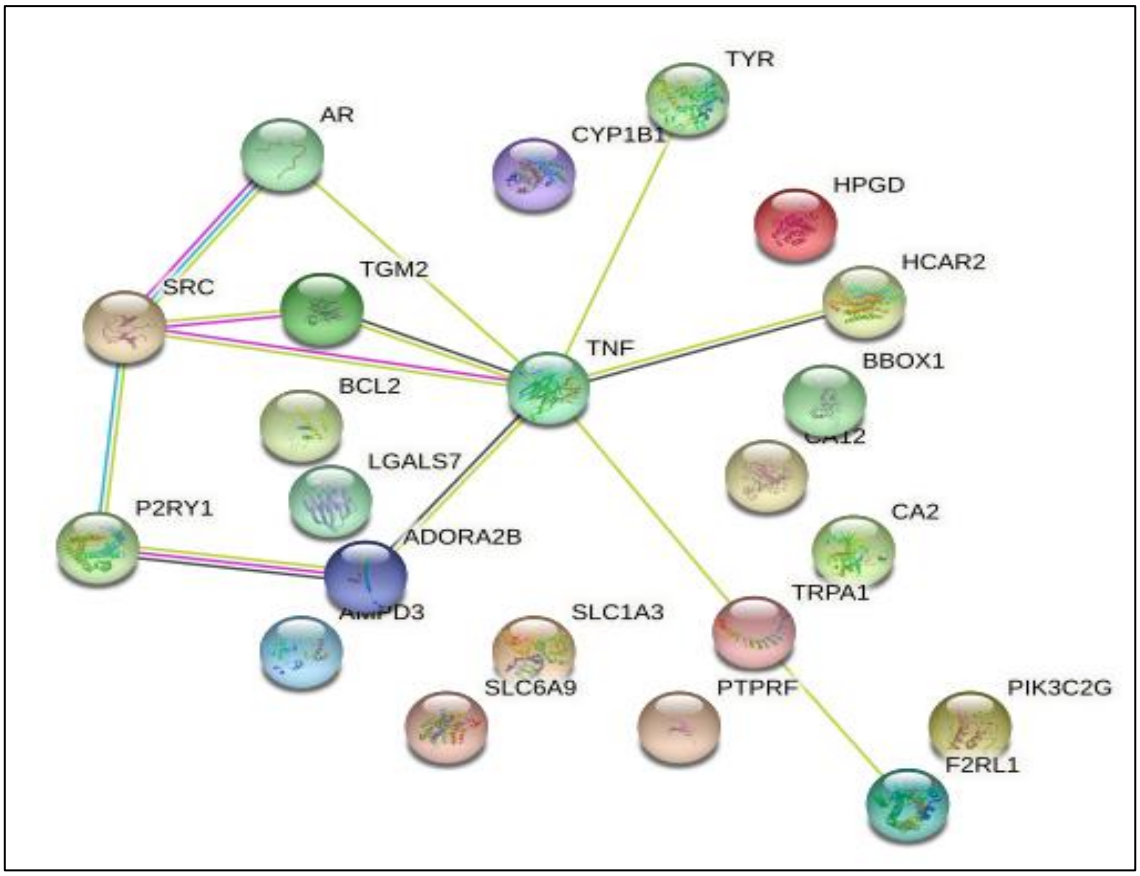


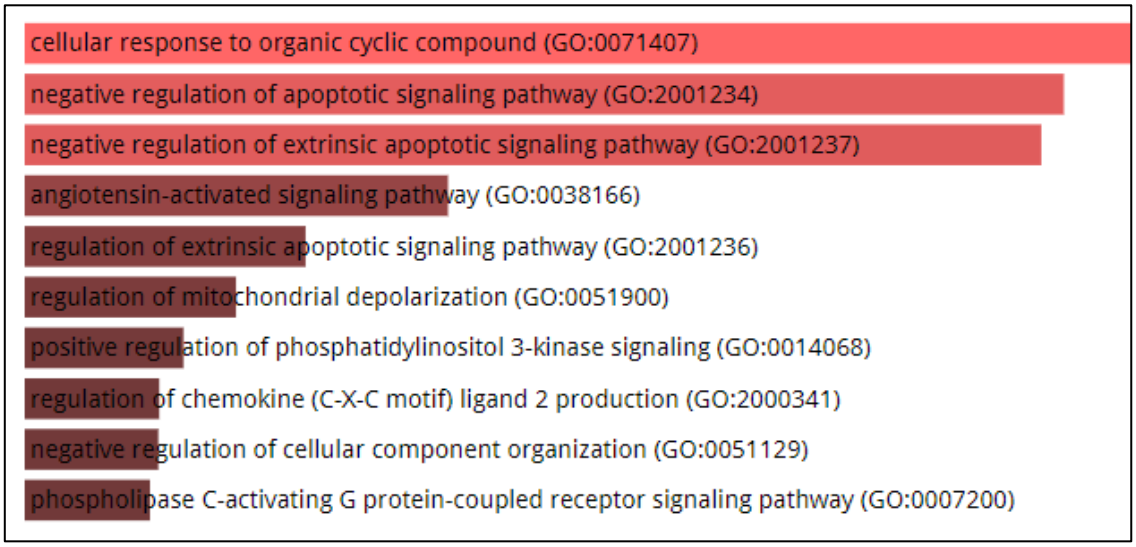
Figure 4: PPI network with the matched DEGs.

4.4 Gene Enrichment Analysis

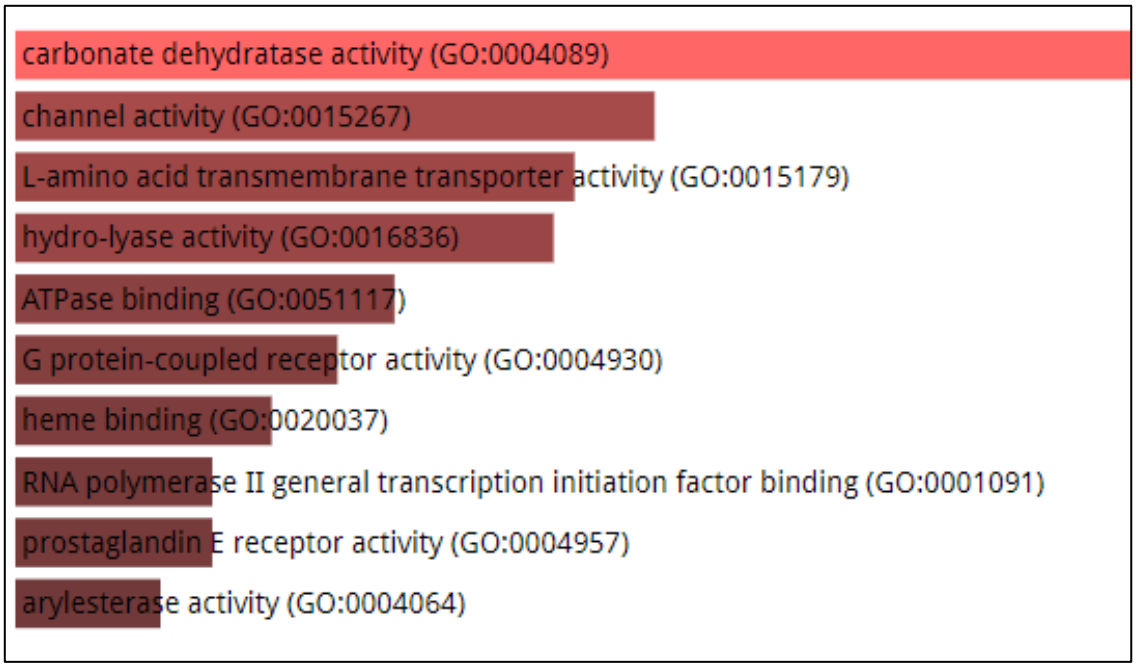
The Enrichr tool was used to perform the gene enrichment analysis. The process was performed on the matching 22 genes. The enrichment analysis was performed in biological process, cellular component, molecular function and pathways affected. A Cytoscape plug-in CLUEGO can also be used for the functional enrichment analysis. For our 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 22 genes, biological processes were over-represented in the Cellular Response to Organic Cyclic Compound for TGM2, and were less expressed in Negative Regulation of Extrinsic Apoptotic Signaling Pathway, Angiotensin-Activated Signaling Pathway, Regulation of Extrinsic Apoptotic Signaling Pathway, Regulation of Mitochondrial Depolarization. The enriched molecular functions were Carbonate dehydratase activity, channel activity, L-amino acid transmembrane transported activity,

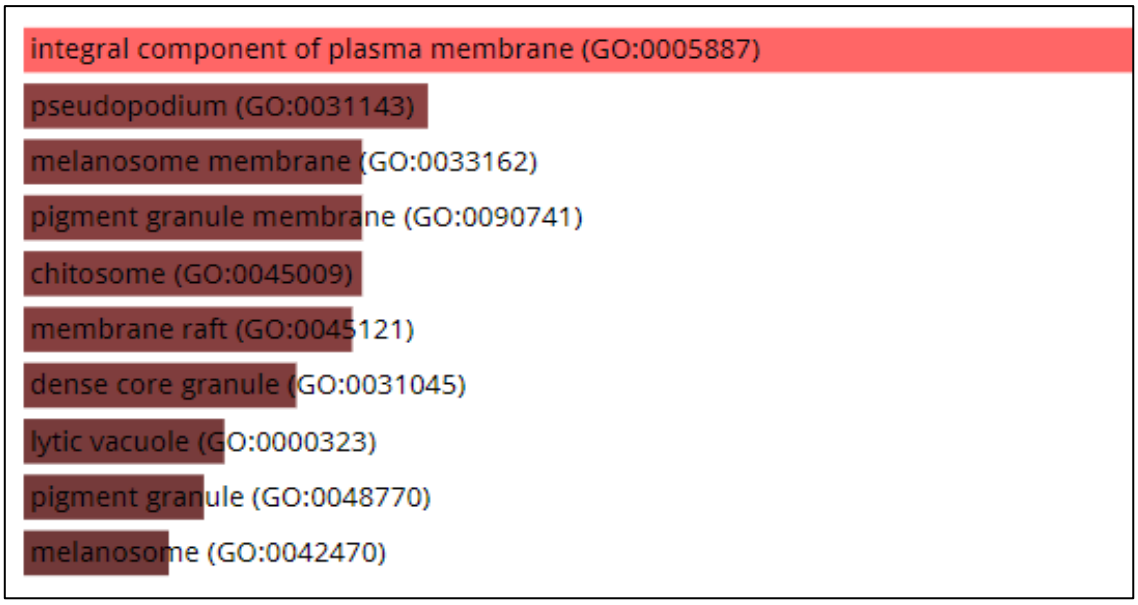
hydro-lyase activity, ATPase binding. Finally, wikipathways found significantly enriched pathways like Aryl Hydrocarbon Receptor Netpath, Nucleotide GPCRs, Oxidative Damage, Insulin signaling. The enrichment result with p value <=0.05 are shown in Figure 5A to 5D.



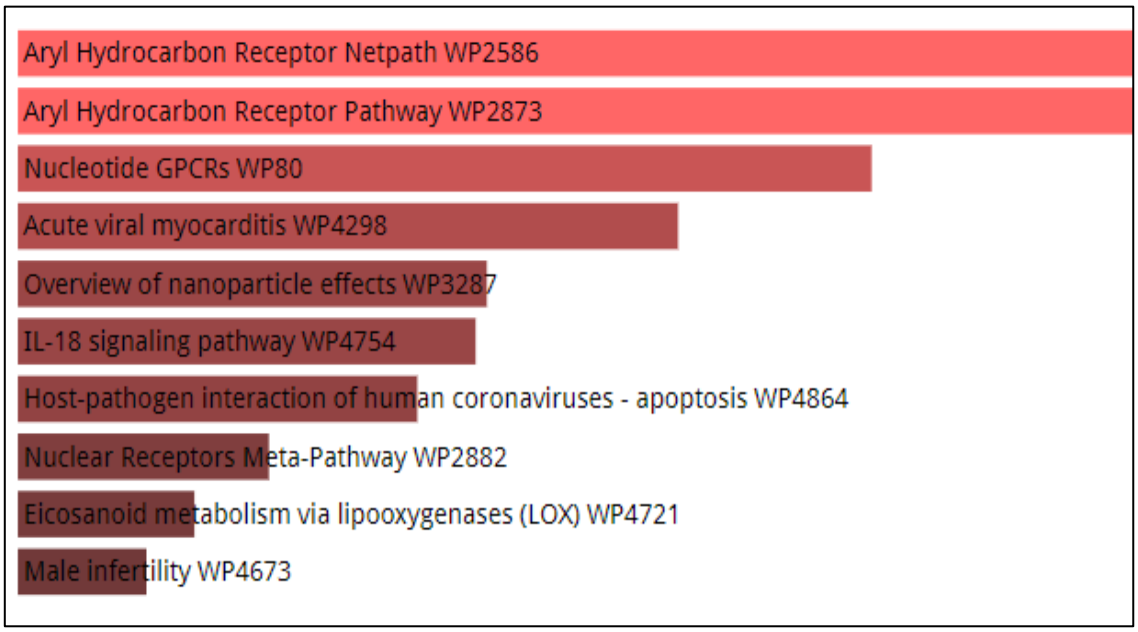
(A)



(B)



(C)



(D)

Figure 5: 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. **(A-D)**

Thus, after analysing the top 5 pathways and their related biological processes the genes that are getting upregulated or overexpressed and downregulated can be found. It is concluded from the result that most of the genes are getting downregulated except

TGM2 which is getting overexpressed. Table 5 shows the brief description about the expression of the genes their functional enrichment.

Table 4: Pathway and GO analysis of down and up-regulated DEGs for T2D.

Term	ID	Description	Adjusted P-Value	Genes
Pathway	WP2586	Aryl Hydrocarbon Receptor Netpath	0.0013	SRC; CYP1B1; TNF
	WP80	Nucleotide GPCRs	0.0033	ADORA2B; P2RY1
	WP4754	IL-18 Signalling Pathway	0.0053	HCAR2; BCL2; TNF; TGM2
	WP3941	Oxidative Damage	0.0085	BCL2; TNF
	WP481	Insulin Signalling	0.0525	PIK3C2G; PTPRF
Biological Process	GO:0071407	Cellular Response to Organic Cyclic Compound	3.91E-04	AR; P2RY1; CYP1B1; TNF; TGM2
	GO:2001237	Negative Regulation of Extrinsic Apoptotic Signalling Pathway	4.06E-04	AR; SRC; BCL2; TNF
	GO:0038166	Angiotensin-Activated Signalling Pathway	0.0044	CA2; SRC
	GO:2001236	Regulation of Extrinsic Apoptotic Signalling Pathway	0.0068	AR; SRC; TNF
	GO:0051900	Regulation of Mitochondrial Depolarisation	0.0069	SRC; BCL2

** Note: Red colour indicates up-regulated and green colour represents down-regulated genes.

CHAPTER 5

DISCUSSION

Diabetes mellitus (DM) is considered as one of the critical illness world-wide as it has a deep negative effect on the quality of life, economic status and mental health of the affected people. Constant regulation related to physical activity, intake of proper diet, is required. In present, proper diagnosis and treatment of diabetes mellitus is still not satisfactory, and the risk percentage of people getting diabetes is still rising. Studies and intensive 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-diabetes medicines, their certain side effects and shortcomings related to durability limit their implications. Understanding the reason for dysfunction of the β-cells of islet of Langerhans at a molecular level can provide better treatment and identification of the pinpointing markers. Some studies found out the marker genes in case of childhood diabetes from one of the GEO data-set by undergoing the 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 data as it provides more perfect results in both qualitative and quantitative when compared to microarray. 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 plant whose 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 nutraceuticals and phytomedicines have comparatively lower side effects thus can be a good alternative for the drug related complications. Traditionally, using of plant derived medicines,

ointments was considered as the best method globally and these approaches has also taken up the credit for its potential as antidiabetic medicine. Hence, the study done on phytochemicals can be considered as a promising way for developing drugs with less complications leading to diabetes management.

By triggering in-silico analysis we were able to find the medicinal plants having an effect for diabetes also upon further target finding could get the phytochemical associated genes. 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, Acacia arabica* (basically stimulates the islets to produce insulin), *Allium sativum* (stabilize blood sugar level and improves blood sugar management), *Aloe vera* (helps in lowering the blood glucose level and also improves the action of tissues towards insulin), *Trigonella foenum-graecum* (lowers the blood sugar level by decreasing the rate of metabolism and carbohydrate uptake) plays an important role in diabetes. Also, its reported that Fenugreek is among the top medicinal plant in terms of safety and efficacy. Although research shows that numerous herbs have the potential of antidiabetic activity as they can modify glucose homeostasis and helps in managing insulin levels. Taking this into account further searched for the plant phytochemicals and their associated genes. It was found that fenugreek had the maximum number 22 (AR, TNF, BBOX1, TYR, CA2, CA12, TGM2, ADORA2B, SLC1A3, HCAR2, CYP1B1, TRPA1, PTPRF, AMPD3, HPGD, P2RY1, LGALS7, BCL2, SRC, PIK3C2G, F2RL1, SLC6A9) matched genes. Already research in the past years has already shown the potential of fenugreek seeds of decreasing blood glucose levels and improves the percentage of glucose tolerance in patients having diabetes. Thus, these 22 genes were taken for further work. A PPI networking was done to find the inner relation between the genes, keeping the confidence score '0.4' there was a 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 DEGs isolated were taken for gene enrichment analysis. In biological process, positive regulation of cellular response to organic compound, Negative Regulation of Extrinsic Apoptotic Signalling Pathway, Angiotensin-Activated Signalling Pathway, Regulation of Extrinsic Apoptotic Signalling Pathway, Regulation of Mitochondrial Depolarisation were found to be most enriched. 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 Aryl hydrocarbon receptor pathway as the highly enriched pathway 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 glucose 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 fenugreek helped in lowering blood glucose level. Similarly, in our study we have also found that genes from various molecules/compounds of fenugreek plant having antidiabetic activity. One of the found phytochemical 4-hydroxyisoleucine associated with several genes like (BBOX1, CA2, CA12) has insulin like actions, Apigenin a potent phytochemical having associated genes (CYP1B1, CA2, CA12, TYR, AR, SRC) helps in lowering the risk of diabetes by stimulating glucose metabolism and also improves the insulin release, folic acid plays an active role against diabetes by lowering fasting insulin levels by regulating the genes like BCL2, PTPRF, ADORA2B, TNF CA2, AMPD3 through the pathways of oxidative damage, 3,4,7-trimethylcoumarin also known as trigoforin shows antidiabetic property by stimulating insulin secretion through biological process of Cellular Response to Organic Cyclic Compound by differential expression of CYP1B1,TGM2 genes where CYP1B1 gets downregulated and TGM2 gets upregulated. Polysterols, an important part of plant cell membrane is helpful against diabetes as in a study it has been seen that supplementation with polysterols (PS) resulted in increase of Fasting Blood Sugar levels,[53] similarly in our study results it can be seen that associated genes of PS specially TNF whose down-regulation is processed by the extrinsic apoptotic signalling pathway contributes to improvement of insulin response and IRS-1 phosphorylation. 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

The chronic disorder like diabetes specific competitive genes and medicinal plant-based phytochemicals were identified by in-silico 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 diabetes. 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 diabetes. 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 as they improve the sensitivity for insulin in the body. Recently, plant-based therapeutics is getting light as several evidences are found for its potential to develop drugs with lesser side-effects. In future more resource and experiments are needed for the unexplored plants to find their metabolites and effects in developing drugs before going for clinical trials just to ascertain any side effects.

CHAPTER 7

REFERENCE

1. Diabetes Fact sheet N°312". WHO. October 2013. Archived from the original on 26 August 2013. Retrieved 25 March 2014.

2. The top 10 causes of death". www.who.int. Retrieved 18 May 2020.

3. Shoback DG, Gardner D, eds. (2011). "Chapter 17". Greenspan's basic & clinical endocrinology (9th ed.). New York: McGraw-Hill Medical. ISBN 978-0-07-162243-1.

4. Kayal, Rajarshi & Kayal, Saheli & Banerjee, Sudip. (2017). Oxidative stress and free radicals related to diabetes: A review. 2. 2455-6548.

5. https://www.idf.org/aboutdiabetes/type-1-diabetes.html.

6. 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 ID 9602016, 10 pages, 2020. https://doi.org/10.1155/2020/9602016.

7. 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

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

9. 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). Antidiabetic Phytochemicals From Medicinal Plants: Prospective Candidates for New Drug Discovery and Development. Frontiers in Endocrinology. 13. 10.3389/fendo.2022.800714.

10. https://doi.org/10.3389/fgene.2017.00174.

11. A. Oyagbemi, M. Salihu, O. Oguntibeju, A. Esterhuyse, and E. E.O.Farombi, "Some Selected Medicinal Plants with Antidiabetic Potentials", in Antioxidant-Antidiabetic Agents and Human Health. London, United Kingdom: IntechOpen, 2014 [Online]. Available: https://www.intechopen.com/chapters/45886 doi: 10.5772/57230.

12. Anjali D Deshpande, Marcie Harris-Hayes, Mario Schootman, Epidemiology of Diabetes and Diabetes-Related Complications, Physical Therapy, Volume 88, Issue 11, 1 November 2008, Pages 1254–1264, https://doi.org/10.2522/ptj.20080020.

13. Ramachandran A. Know the signs and symptoms of diabetes. Indian J Med Res. 2014 Nov;140(5):579-81. PMID: 25579136; PMCID: PMC4311308.

14. Matough FA, Budin SB, Hamid ZA, Alwahaibi N, Mohamed J. The role of oxidative stress and antioxidants in diabetic complications. Sultan Qaboos Univ Med J. 2012 Feb;12(1):5-18. doi: 10.12816/0003082. Epub 2012 Feb 7. PMID: 22375253; PMCID: PMC3286717.

15. Mittal M, Siddiqui MR, Tran K, Reddy SP, Malik AB. Reactive oxygen species in inflammation and tissue injury. Antioxid Redox Signal. 2014 Mar 1;20(7):1126-67. doi: 10.1089/ars.2012.5149. Epub 2013 Oct 22. PMID: 23991888; PMCID: PMC3929010.

16. McMillin JM. Blood Glucose. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition. Boston: Butterworths; 1990. Chapter 141.

17. Mechanisms of Insulin Action and Insulin Resistance Max C. Petersen and Gerald I. Shulman Physiological Reviews 2018 98:4, 2133-2223.

18. Ndisang JF, Vannacci A, Rastogi S. Insulin Resistance, Type 1 and Type 2 Diabetes, and Related Complications 2017. J Diabetes Res. 2017;2017:1478294. doi: 10.1155/2017/1478294. Epub 2017 Nov 15. PMID: 29279853; PMCID: PMC5723935.

19. Ang GY. Age of onset of diabetes and all-cause mortality. World J Diabetes. 2020 Apr 15;11(4):95-99. doi: 10.4239/wjd.v11.i4.95. PMID: 32313608; PMCID: PMC7156298.

20. Kalra S, Mukherjee JJ, Venkataraman S, Bantwal G, Shaikh S, Saboo B, Das AK, Ramachandran A. Hypoglycemia: The neglected complication. Indian J Endocrinol Metab. 2013 Sep;17(5):819-34. doi: 10.4103/2230-8210.117219. PMID: 24083163; PMCID: PMC3784865.

21. https://www.diabetes.org/diabetes/genetics-diabetes.

22. Cui Y., Li G., Li S., Wu R. (2010) Designs for Linkage Analysis and Association Studies of Complex Diseases. In: Bang H., Zhou X., van Epps H., Mazumdar M. (eds) Statistical Methods in Molecular Biology. Methods in Molecular Biology (Methods and Protocols), vol 620. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-60761-580-4_6.

23. Morwessel NJ. The genetic basis of diabetes mellitus. AACN Clin Issues. 1998 Nov;9(4):539-54. doi: 10.1097/00044067-199811000-00009. PMID: 9855864.

24. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, Sjögren M, Ling C, Eriksson KF, Lethagen AL, Mancarella R, Berglund G, Tuomi T, Nilsson P, Del Prato S, Groop L. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007 Aug;117(8):2155-63. doi: 10.1172/JCI30706. PMID: 17671651; PMCID: PMC1934596.

25. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002 May;3(5):391-7. doi: 10.1038/nrg796. PMID: 11988764.

26. McCarthy, M.I., Zeggini, E. Genome-wide association studies in type 2 diabetes. Curr Diab Rep 9, 164–171 (2009). https://doi.org/10.1007/s11892-009-0027-4.

27. Ip W, Chiang YT, Jin T. The involvement of the wnt signaling pathway and TCF7L2 in diabetes mellitus: The current understanding, dispute, and perspective. Cell Biosci. 2012 Aug 14;2(1):28. doi: 10.1186/2045-3701-2-28. PMID: 22892353; PMCID: PMC3468386.

28. Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, Scott LJ, Narisu N, Chines PS, Skol A, Stringham HM, Petrie J, Erdos MR, Swift AJ, Enloe ST, Sprau AG, Smith E, Tong M, Doheny KF, Pugh EW, Watanabe RM, Buchanan TA, Valle TT, Bergman RN, Tuomilehto J, Mohlke KL, Collins FS, Boehnke M. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. Diabetes. 2007 Jan;56(1):256-64. doi: 10.2337/db06-0461. PMID: 17192490.

29. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. Curr Diab Rep. 2009 Apr;9(2):164-71. doi: 10.1007/s11892-009-0027-4. PMID: 19323962; PMCID: PMC2694564.

30. Ling C, Groop L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes. 2009 Dec;58(12):2718-25. doi: 10.2337/db09-1003. PMID: 19940235; PMCID: PMC2780862.

31. Amiri-Dashatan N, Koushki M, Abbaszadeh HA, Rostami-Nejad M, Rezaei-Tavirani M. Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry. Iran J Pharm Res. 2018 Fall;17(4):1523-1536. PMID: 30568709; PMCID: PMC6269565.

32. Kong M, Xie K, Lv M, Li J, Yao J, Yan K, Wu X, Xu Y, Ye D. Anti-inflammatory phytochemicals for the treatment of diabetes and its complications: Lessons learned and future promise. Biomed Pharmacother. 2021 Jan;133:110975. doi: 10.1016/j.biopha.2020.110975. Epub 2020 Nov 16. PMID: 33212375.

33. B. Daramola, G.O. Adegoke, Chapter 25 - Bitter Kola (Garcinia kola) Seeds and Health Management Potential, Editor(s): Victor R. Preedy, Ronald Ross Watson, Vinood B. Patel, Nuts and Seeds in Health and Disease Prevention, Academic Press, 2011, Pages 213-220, ISBN 9780123756886, https://doi.org/10.1016/B978-0-12-375688-6.10025-8.

34. K.V. Peter, M.R. Shylaja, 1 - Introduction to herbs and spices: definitions, trade and applications, Editor(s): K.V. Peter, In Woodhead Publishing Series in Food Science, Technology and Nutrition, Handbook of Herbs and Spices (Second Edition), Woodhead Publishing, 2012, Pages 1-24, ISBN 9780857090393, https://doi.org/10.1533/9780857095671.1.

35. Firdous SM. Phytochemicals for treatment of diabetes. EXCLI J. 2014 May 6;13:451-3. PMID: 26417272; PMCID: PMC4464495.

36. Bacanli M, Dilsiz SA, Başaran N, Başaran AA. Effects of phytochemicals against diabetes. Adv Food Nutr Res. 2019;89:209-238. doi: 10.1016/bs.afnr.2019.02.006. Epub 2019 Mar 4. PMID: 31351526.

37. 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 ID 9602016, 10 pages, 2020. https://doi.org/10.1155/2020/9602016.

38. 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.

39. 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.

40. Grigoriev A. A relationship between gene expression and protein interactions on the proteome 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.

41. 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.

42. 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.

43. 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.

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

45. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41(Database issue):D991-D995. doi:10.1093/nar/gks1193.

46. Wimmer RA, Leopoldi A, Aichinger M, Wick N, Hantusch B, Novatchkova M, Taubenschmid J, Hämmerle M, Esk C, Bagley JA, Lindenhofer D, Chen G, Boehm M, Agu CA, Yang F, Fu B, Zuber J, Knoblich JA, Kerjaschki D, Penninger JM. Human blood vessel organoids as a model of diabetic vasculopathy. Nature. 2019 Jan;565(7740):505-510. doi: 10.1038/s41586-018-0858-8. Epub 2019 Jan 16. PMID: 30651639; PMCID: PMC7116578.

47. 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.

48. 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 12;8(1):4329. doi: 10.1038/s41598-018-22631-z. PMID: 29531263; PMCID: PMC5847565.

49. Patel DK, Prasad SK, Kumar R, Hemalatha S. An overview on antidiabetic medicinal plants having insulin mimetic property. Asian Pac J Trop Biomed. 2012 Apr;2(4):320-30. doi: 10.1016/S2221-1691(12)60032-X. PMID: 23569923; PMCID: PMC3609288.

50. 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.

51. 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 2016 Oct 18. PMID: 27924014; PMCID: PMC5210637.

52. 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.

53. 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 levels in 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.