

Integrating RNA-Seq analysis with machine learning approaches for the identification of hub genes affected by exposure of Bisphenol-A in patients suffering from Type 2 Diabetes

A THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF TECHNOLOGY in BIOINFORMATICS

Submitted by
WAYEZ NAQVI
(24/BIO/15)

Under the Supervision of
Prof. YASHA HASIJA



DEPARTMENT OF BIOTECHNOLOGY

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

May, 2026

ACKNOWLEDGEMENT

First and foremost, I express my heartfelt gratitude to the Almighty for granting me the strength, health, and determination to complete this research work.

I extend my sincere thanks to my supervisor, **Prof. Yasha Hasija**, whose guidance, support, and insightful feedback were instrumental throughout the course of this thesis. Her constant encouragement and deep knowledge in the field of bioinformatics have helped shape this work and my academic journey.

I am also thankful to the **Department of Biotechnology**, Delhi Technological University, for providing me with the infrastructure, resources, and academic environment necessary to carry out this research.

I would like to express my appreciation to all the faculty members and technical staff of the department for their valuable assistance and cooperation whenever needed.

A special thanks to my peers and friends for their continuous motivation, stimulating discussions, and moral support throughout the course of my study. Their presence made this journey more enriching and enjoyable.

Finally, I am deeply indebted to my parents and family for their unwavering support, patience and encouragement during every stage of this work. Their sacrifices and belief in me have been the greatest source of strength.

This thesis would not have been possible without the contribution and support of all the above-mentioned individuals. I remain grateful to each and every one of them.

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Wayez Naqvi, hereby declare that the work which is being presented in the thesis entitled **“Integrating RNA-Seq analysis with machine learning approaches for the identification of hub genes affected by exposure of Bisphenol A in patients suffering from Type 2 Diabetes”** in partial fulfilment of the requirements for the award of the Degree of Master of Technology, submitted in the **Department of Biotechnology**, Delhi Technological University is an authentic record of my own work carried out during the period from **January 2026** to **May 2026** under the supervision of **Prof. Yasha Hasija**. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Vm
29.05.26

Signature of Supervisor

Wayez

Candidates' Signature

Date: 29.05.2026

DEPARTMENT OF BIOTECHNOLOGY**DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE BY THE SUPERVISOR

Certified that **Wayez Naqvi** (Roll No. 24/BIO/15) has carried out their research work presented in this thesis entitled **“Integrating RNA-Seq analysis with machine learning approaches for the identification of hub genes affected by exposure of Bisphenol-A in patients suffering from Type 2 Diabetes”** for the award of **Master of Technology** from Department of Biotechnology, Delhi Technological University, under my supervision. The thesis embodies results of original work and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University Institution.

WM
29.05.26

Prof. Yasha Hasija

Head of the Department & Supervisor

Delhi Technological University

Date: 29.05.2026

ABSTRACT

Type 2 Diabetes (T2D) is a metabolic disorder caused (Butler, Janson and Bonner-Weir) due to development of insensitivity of insulin (Butler, Janson and Bonner-Weir) and loss of β -cell functioning (Butler, Janson and Bonner-Weir) (Iqra Hameed) (Vito.) (Rorsmam.). Evidence suggests that exposure to bisphenol-A (BPA) may contribute in the development and progression of type 2 diabetes. Thus, this study integrates bioinformatics and machine learning approaches to identify key genes and pathways involved linking BPA with T2DM. Datasets related to T2D were retrieved from GEO, which were then pre-processed, normalized and batch corrected. DEG analysis was then done to identify the up and down regulating genes. Following of which the data was split into training and test set. BPA related genes were then retrieved from database comparative toxicogenomics database, swiss target prediction and STITCH. While, disease related genes were retrieved from GeneCards, MalaCards and OMIM. Results from these were then overlapped to identify hub genes which were assessed using machine learning approaches RF, SVM and LASSO to identify potential candidate biomarkers. Next, ROC curve analysis was done on the genes i.e ABCG1, AMPD2 and HBA2. This integrated approach also did pathway enrichment of the potential diagnostic biomarkers which highlighted pathways related to metabolic regulation, mitochondrial function and inflammatory responses. Lastly, molecular docking was used for assessing the binding interaction occurring between BPA and ABCG1, AMPD2 and HBA2 which yielded a binding affinity of -7.4, -8.1 and – 8.1 kcal/mol respectively. Overall this study helped in identifying a potential link between BPA and T2DM and in identifying the potential candidates that may contribute progression of disease.

Keywords: Type 2 Diabetes, Bisphenol-A, DEG, Machine Learning, Molecular Docking.

TABLE OF CONTENT

ACKNOWLEDGEMENT	ii
CANDIDATE’S DECLARATION	iii
CERTIFICATE BY THE SUPERVISOR	iv
ABSTRACT	v
TABLE OF CONTENT	vi - vii
LIST OF TABLES	viii
LIST OF FIGURES	ix - xii
LIST OF ABBREVIATIONS	xii
1 INTRODUCTION	1 - 2
2 LITERATURE REVIEW	3 - 6
3 MATERIALS & METHODS	7 - 21
3.1 Tools used	7 - 18
3.1.1 NCBI.....	7
3.1.2 GEO.....	7 - 8
3.1.3 GeneCards.....	8 - 9
3.1.4 MalaCards	9
3.1.5 OMIM.....	10
3.1.6 PubChem	10 - 11
3.1.7 CTD.....	11 - 12
3.1.8 STP.....	12 - 13
3.1.9 STITCH.....	13 - 15
3.1.10 UniProtKB.....	14 - 15
3.1.11 RCSB PDB.....	15 - 16
3.1.12 Discovery Studio Visualizer.....	16
3.1.13 Rstudio	16 - 17
3.1.14 AutoDock Vina	17
3.1.15 GTEEx	18
3.2 Methodology	19 - 21
3.2.1 Dataset Retrieval	19
3.2.2 Data Preprocessing.....	19 - 20
3.2.3 Differential Gene Expression Analysis.....	20
3.2.4 Retrieval of BPA targets and T2DM genes.....	20
3.2.5 Functional annotation.....	20
3.2.6 Identification of candidate genes.....	20
3.2.7 Training and Validation set construction	20 - 21
3.2.8 Machine learning analysis.....	21
3.2.9 Diagnostic validation using ROC curve.....	21
3.2.10 Gene set enrichment analysis	21

3.2.11 Tissue expression analysis	21
3.2.12 Molecular docking.....	21
4 RESULTS & DISCUSSION	22 - 40
4.1 Dataset Retrieval.....	22
4.2 Data Preprocessing	23 - 24
4.3 Differential Gene Expression Analysis.....	24 - 25
4.4 Retrieval of BPA targets and T2DM genes.....	25 - 26
4.5 Functional annotation	26 - 28
4.6 Identification of candidate genes	29
4.7 Training and Validation set construction.....	30
4.8 Machine learning based hub gene identification	30 - 34
4.9 ROC curve analysis and diagnostic validation	35 - 36
4.10 Gene set enrichment analysis.....	36 - 37
4.11 Tissue expression analysis.....	38
4.12 Molecular docking	39 - 40
5 CONCLUSION	41 - 42
6 REREFENCES	43 – 58
7 PLAGIARISM VERIFICATION	59

LIST OF TABLES

Table No.	Title	Page No.
1	Summary of the selected datasets available on GEO with their respective Bioproject and SRA accession ID's, and sample sizes.	27
2	Table showing the 52 genes obtained after the overlap of the DEG and WGCNA genes.	34
3	Table showing the set genes obtained through the use of machine learning approaches.	37

LIST OF FIGURES

Figure No.	Title	Page No.
1	Homepage of NCBI	8
2	Homepage of GEO	9
3	Homepage of GeneCards.	10
4	Homepage of MalaCards.	11
5	Homepage of OMIM	12
6	Homepage of PubChem.	14
7	Homepage of Comparative Toxicogenomics Database (CTD)	15
8	Homepage of Swiss Target Prediction (STP) Database.	17
9	Homepage of STITCH	18
10	Homepage of UniProtKB	20
11	Homepage of RCSB PDB	21
12	Homepage of GTEx	23
13	Flowchart depicting the pipeline followed in this study	24
14	Initial PCA plot of the T2D datasets. Here DS-1 represents samples of GSE154881, DS-2 represents samples of GSE153315, DS-3 represents samples of GSE184050, and DS-4 represents samples of GSE114192	28

15	PCA plot after batch correction. Here DS-1 represents samples of GSE154881, DS-2 represents samples of GSE153315, DS-3 represents samples of GSE184050, and DS-4 represents samples of GSE114192	29
16	Volcano plot depicting the differentially expressed genes obtained through DESeq2 (Wayez Naqvi). Where x-axis represents the log ₂ (fold change) value and y-axis represent the -log ₁₀ (adjusted p value) (Wayez Naqvi). Blue coloured dot represents the down-regulated genes, red coloured dot represents up-regulated genes while grey coloured dot represents the non-significant genes (Wayez Naqvi)	30
17	Venn diagram showing the overlapping genes between BPA target genes and T2DM disease genes	31
18	Functional profiling of the intersecting BPA T2D genes using KEGG terms	32
19	Functional profiling of the intersecting BPA T2D genes using Gene ontology terms	33
20	Feature importance curve of the top 10 Random forest genes	36
21	Out of bag error estimation of the Random Forest model	36
22	Feature selection performance of the SVM-RFE using 5-fold cross-validation	37
23	10-fold cross-validation error curve of LASSO regression analysis	37
24	Coefficient trajectory plot of LASSO model, where each coloured lines represents the coefficient of each candidate genes taken in consideration.	38
25	Venn diagram showing overlapping genes obtained from the three machine learning approaches	38

26	Figure showing the 3 gene ROC curve	40
27	Figure showing the jittered boxplot and single gene ROC curve of (A) ABCG1; (B) AMPD2; and (C) HBA2	41
28	Panelled image depicting the top significantly enriched GO-GSEA pathways (A&B); and top significantly enriched KEGG GSEA pathways (C&D)	42
29	Figure depicting gene expression analysis of (A) ABCG1; (B) AMPD2; and (C) HBA2 across all the tissues	43
30	Complex formed between Bisphenol A and (A) Adenosine monophosphate deaminase 2; (B) ATP Binding Cassette sub family G member 1	44
31	Complex formed between Bisphenol A and (C) Haemoglobin Subunit Alpha 2	45

LIST OF ABBREVIATIONS

Abbreviation	Full Form
BPA	Bisphenol A
T2D/T2DM	Type 2 Diabetes/Type 2 Diabetes Mellitus
DEG	Differentially Expressed Gene(s)
GEO	Gene Expression Omnibus
NCBI	National Center for Biotechnology Information
CTD	Comparative Toxicogenomics Database
STP	SwissTargetPrediction
STITCH	Search Tool for Interacting Chemicals
OMIM	Online Mendelian Inheritance in Man
PDB	Protein Data Bank
GTE _x	Genotype-Tissue Expression
RF	Random Forest
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine Recursive Feature Elimination
LASSO	Least Absolute Shrinkage and Selection Operator
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
PCA	Principal Component Analysis
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
GSIS	Glucose-Stimulated Insulin Secretion
EDC	Endocrine Disrupting Chemical
ER α	Estrogen Receptor Alpha
ER β	Estrogen Receptor Beta
GPR30	G Protein-Coupled Receptor 30
IL-6	Interleukin-6
MCP-1	Monocyte Chemoattractant Protein-1
ROS	Reactive Oxygen Species
PI3K	Phosphoinositide 3-Kinase
Akt	Protein Kinase B
GLUT4	Glucose Transporter Type 4
PPAR γ	Peroxisome Proliferator-Activated Receptor Gamma
C/EBP α	CCAAT/Enhancer Binding Protein Alpha
NF- κ B	Nuclear Factor Kappa B
MAPK	Mitogen-Activated Protein Kinase
RNA-seq	RNA Sequencing
SNP	Single Nucleotide Polymorphism
EMBL-EBI	European Molecular Biology Laboratory – European Bioinformatics Institute
UniProtKB	Universal Protein Knowledgebase
FP2	Fingerprint Version 2
ES5D	ElectroShape 5-Dimensional Descriptor
SMILES	Simplified Molecular Input Line Entry System
TTD	Therapeutic Target Database
GLIDA	GPCR Ligand Database
CRP	C-Reactive Protein
TNF- α	Tumor Necrosis Factor Alpha

INTRODUCTION

Type 2 Diabetes (T2D) is a metabolic disorder caused (Butler, Janson and Bonner-Weir) due to development of insensitivity of insulin (Unai Galicia-Garcia) (Butler, Janson and Bonner-Weir) and loss of β -cell functioning (Unai Galicia-Garcia) (Butler, Janson and Bonner-Weir) (Iqra Hameed) (Vito.) (Rorsmam.). Hyperglycaemia, dyslipidaemia, low-grade inflammation (Anne M Minihane), and up to some extent gut microbial dysbiosis (Patricia M. Bock) play a role in development of type 2 diabetes (Vito.). All these factors contribute to the loss of beta cell functionality and development of insulin resistance, the major hallmarks of type 2 diabetes (Vito.). As discussed before, hyperglycaemia induces insensitivity to insulin and β cell dysfunction (Vito.), as they progress further, they fuel hyperglycaemia further ultimately promoting type 2 diabetes, therefore forming a vicious cycle (Cerf.).

Bisphenol A (BPA) a prominent EDC is found in various sources like epoxy resins, lining of food cans, recycled paper and coatings of CD and DVD (Rubin.) (La-or Chailurkit). Since it is used in food packaging and in other day to day products, significant amount of human exposure occurs through the oral route (Francesca Farrugia). BPA acts as a xenoestrogen and binds to the extranuclear receptors $ER\alpha$ and $Er\beta$ (A. B. Paloma Alonso-Magdalena) (Sergi Soriano). The class of xenoestrogens are also known as foreign estrogens (Xiaoqiang Wang) due to the fact that they exhibit estrogen like behaviour and despite being produced exogenously i.e outside the body (Xiaoqiang Wang). By behaviour we mean that xenoestrogen attaches itself to the extranuclear receptors and govern the expression of associated genes (Xiaoqiang Wang). Apart from activating $ER\alpha$ and $Er\beta$ (Xiaoqiang Wang).. Xenoestrogen can also stimulate the membrane bound ER (GPER/GPR30) (Pelin Yaşar) (Xiaoqiang Wang). Through literature study it was identified that BPA induced stimulation of GPR30 allows for more activation of the ERK signalling pathways (Wolfgang Völkel) (Dong.). BPA has also been found to suppress production and release of adiponectin, which leads to low level of adiponectin in the body (Francesca Farrugia).

Pregnancy is often associated with gestational diabetes as the body fails to produce adequate amounts of insulin to counter shield insulin resistance (E. V. Paloma Alonso-Magdalena) (M. G.-A. Paloma Alonso-Magdalena). Glucose stimulated insulin secretion or GSIS serves a prominent role by keeping a check on glucose levels in the body (Sergi Soriano). By exhibiting estrogen like behaviour, BPA acts on $Er\alpha$ (A. B. Paloma Alonso-Magdalena) to altering insulin levels and enhancing GSIS activity (Vito.). Mouse model studies further highlighted that (Vito.) BPA induced altered glycaemic response and hyperinsulinemia state along with increasing levels of circulating insulin, a characteristic trait of insulin resistance (Tetsuya Adachi). This ultimately leads to metabolic complications which also increases the risk of Type 2 diabetes (Vito.). This work focuses on linking BPA exposure to T2D, and in pinpointing the exact elements or factors at play and outline the emerging mechanisms involved (Francesca Farrugia).

LITERATURE REVIEW

Arguably one of the most widely used industrial synthetic chemical compounds in the world, bisphenol A (BPA) present in epoxy resins (Harner.). Such products are applied as packaging material for food, reusable drinking containers, liners of canned foods, medical devices, thermally printed invoices (Maria K. Björnsdotter), and a number of other household items. Because of such extensive applications and constant discharges into the environment, BPA contamination of the human population has been considered inevitable. Scientific interest in possible adverse health consequences of BPA has been increasing steadily during the last decades due to its endocrine disruption capabilities and association with chronic diseases, especially T2DM. It is a disorder with a complex set of characteristics that include insulin resistance, insulin dysfunction, hyperglycemia, inflammation, oxidative stress, and disorders of lipid metabolism (Unai Galicia-Garcia). Yet recently it has become evident that environmental toxins, among which BPA can be singled out as an example of endocrine disrupting compounds (EDCs), can also serve as significant contributors to diabetes development (Sargis.). First scientific reports showed the harmful influence of BPA on hormonal communication, reproduction, and development (Winn.). But later scientists found out that the negative impact of BPA is much wider – it can greatly change metabolic processes, related to glucose homeostasis and insulin activity. That is why BPA was referred to as a “metabolic disruptor.” (Soheila Rahmani)

One of the first significant epidemiological studies on BPA, was carried out research on NHANES data for the US population (Iain A. Lang). The authors proved that high levels of BPA in urine were significantly related to diabetes, heart diseases, and altered hepatic enzymes (Iain A. Lang). Such research drew much scientific interest because it showed the link between BPA exposure and chronic metabolic diseases at the population level (Iain A. Lang). After such study, several other epidemiological works published around the world supported similar results. Song et al., analyzing NHANES data, found a high correlation between increased BPA content in urine and prevalence of type 2 diabetes mellitus among adults. Similar results were obtained from studies conducted in Chinese, Korean, Japanese, and even European populations. Further meta-analyses have corroborated

these observations. In their systematic review and meta-analysis, researchers analysed epidemiological evidence associating exposure to BPA with cardiometabolic diseases (Fanny Rancière). The authors noted a consistent relationship between BPA exposure and a greater likelihood of developing obesity, insulin resistance, and T2DM (Fanny Rancière). Another meta-analysis was done of cross-sectional and prospective studies which revealed a positive correlation between BPA exposure and T2DM development (Semi Hwang). Despite some conflicting reports related to exposure evaluation methodology, population type, study size, and potential confounders, most evidence indicated a strong association between BPA exposure and metabolic problems (Semi Hwang).

Additional support for this relationship's biological plausibility came from research carried on animals. Since exposure to endocrine-disrupting substances in the environment during pregnancy and children may cause metabolic problems later in life, the significance of this issue has grown. (Jie Wei). In addition, there is evidence that BPA exposure affects the physiology of pancreatic β cells (S. M. Paloma Alonso-Magdalena). Insulin production and secretion from β cells are necessary to maintain glucose balance in the body (S. M. Paloma Alonso-Magdalena). Although acute exposure to BPA stimulates these β cells, prolonged one causes their exhaustion and dysfunction (S. M. Paloma Alonso-Magdalena). Some researchers demonstrated that BPA exposure can affect β cell Ca^{2+} signalling, cause an increase in oxidative stress and lead to apoptosis. It should be noted that β cell dysfunction is a hallmark feature of type 2 diabetes development (Francesca Farrugia). The second mechanism through which BPA may induce T2DM is insulin resistance (Francesca Farrugia). This refers to the phenomenon where the ability of muscle, liver, and fat tissues to respond to insulin becomes compromised (Francesca Farrugia).

Obesity is yet another crucial biological connection between exposure to BPA and T2DM. Exposure to BPA is known to cause obesity owing to its function as obesogen, which means its ability to promote adipogenesis and accumulation of lipids (Mónica García García). It has been proved that BPA exposure causes stimulation of preadipocyte differentiation into adipocytes through transcription factors like $\text{PPAR}\gamma$ and $\text{C/EBP}\alpha$ (Jonathan G Boucher). Formation of more adipocytes results in excess fat

deposition, leading to obesity that further leads to insulin resistance and risk for diabetes. Another study found that BPA increased adipogenesis via transcriptional regulation mechanisms (Jonathan G Boucher). Apart from adipogenesis, BPA exposure has also been shown to lead to changes in lipid metabolism (Jonathan G Boucher). Another paper claimed with their study that low levels of BPA exposure led to the induction of gene expression linked with fat synthesis and triacylglycerol accumulation in liver tissue in mice (Alice Marmugi). BPA-induced oxidative stress was also noted as yet another key factor responsible for metabolic disturbances. Elevated ROS levels affect the normal functioning of proteins, lipids, and nucleic acids, which results in metabolic imbalance (Rosaria Meli). Specifically, pancreatic β -cells appear to be highly susceptible to oxidative stress due to the poor enzymatic antioxidant capacity (Rosaria Meli) (Shehreen Amjad). Disruption of mitochondria function is another important characteristic of BPA-induced metabolic diseases (Y Lin). Mitochondria play an important role in energy metabolism, ATP synthesis, and control of cell's redox status (Y Lin). It was shown that BPA affects the mitochondrial membrane potential, mitochondrial oxidation, and mitochondrial biogenesis.

Inflammation is one of the mechanisms thought to mediate BPA-induced T2DM development. It has been found that BPA upregulates inflammatory cascades like NF- κ B and MAPK pathways resulting in induction of inflammation in adipose tissue, liver, and pancreas (Xu Hong). This chronic inflammation further affects the insulin signalling cascade. There has also been recent interest in the role of epigenetic regulation of BPA-related metabolic disorders (F. J.-B. Paloma Alonso-Magdalena). The dysregulation of miRNAs may have an effect on insulin pathways and β -cell survival, thereby aiding in the development of diabetes (Jin Hee Kim). There is great advancement by high throughput omics in terms of understanding metabolic disorders caused by BPA (Jin Hee Kim). Such pathways include insulin signalling pathways, inflammation pathways, oxidative stress response pathways, lipid metabolism pathways, apoptosis pathways and mitochondrial dysfunction pathways. Recent advancements in bioinformatics and systems biology have led to the identification of biomarkers related to BPA-induced T2DM through various computational approaches. WGCNA facilitates the

recognition of interconnected genes and their role in regulatory modules and biological pathways, which help in developing the pathogenesis of a disease. Various genes and their roles in BPA related diabetes have been recognized using WGCNA. Furthermore, advancements in machine learning have significantly improved biomarker discovery and prediction in the metabolic research related to BPA. These machine learning methods are currently being utilized on large transcriptomic data to find out the hub genes responsible for T2DM development. The combination of WGCNA with machine learning approaches provides more precise identification of biomarkers. Some genes that regulate mitochondrial dysfunction, inflammatory pathway, glucose homeostasis, and lipid metabolism have been proposed as hub genes.

MATERIALS & METHODS

3.1 Tools Used:

3.1.1 NCBI:

NCBI s (Conrad L Schoch) (<https://www.ncbi.nlm.nih.gov/>) houses a series of databases relevant to the field of biotechnology and biomedicine. Major databases include GenBank (Conrad L Schoch), Gene, Nucleotide, Protein, dbSNP, PubMed etc. All these databases can be accessed through the Entrez search engine.

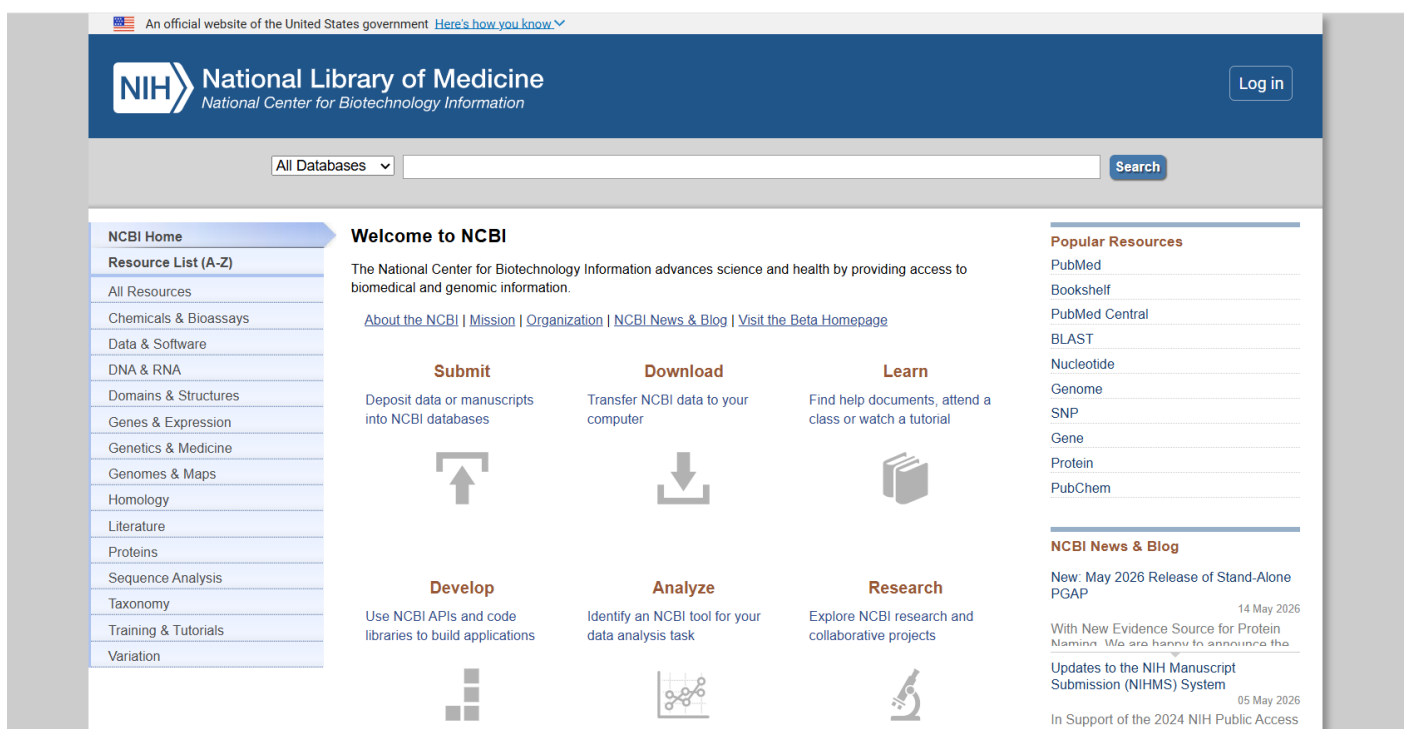


Figure 1: Homepage of NCBI

3.1.2 GEO:

GEO contains (<https://www.ncbi.nlm.nih.gov/geo/>) various types of high-throughput functional genomics data in its publicly accessible database (Tanya Barrett) (Ron Edgar). GEO was designed with the purpose of providing a strong (Tanya Barrett), extensible database in which high-throughput data can be effectively stored and users may easily query, identify, evaluate, and download studies and gene expression profiles of interest (Tanya Barrett) (Ron Edgar). This database enables accessing raw sequencing data from different experiments around the world, thus enabling using raw data from previous studies in subsequent analyses (Ron Edgar).

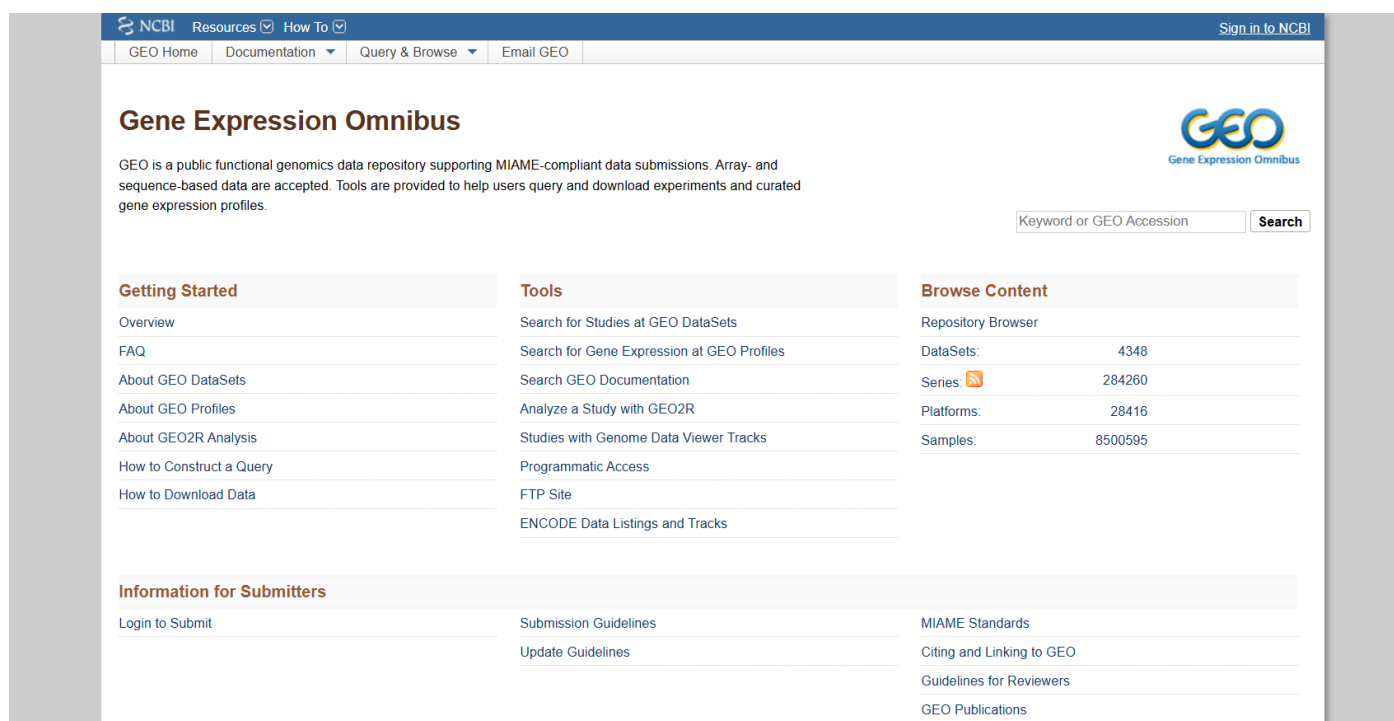


Figure 2: Homepage of GEO

3.1.3 GeneCards:

GeneCards (<https://www.genecards.org/>) currently versioned at 5.26 is a user-friendly database that provides information on all annotated and predicted human genes (N. R.-G. Gil Stelzer). It houses data

on a total of 443,342 genes. The knowledgebase automatically incorporates multi omics level data gene-centric data from myriad number of sources (N. R.-G. Gil Stelzer).



Figure 3: Homepage of GeneCards

3.1.4 MalaCards:

MalaCards (<https://www.malacards.org/>) with current version 5.26; houses information on a total of 22,560 disorders and 15,692 gene-related disorders (Noa Rappaport). It organizes disease-related information into global categories such as cancer, genetic, fetal, infectious, metabolic, and uncommon. It also categorizes disorders based on anatomical categories, such as cardiovascular, endocrine, neurological, and mental problems (Noa Rappaport).

3.1.5 OMIM:

OMIM stands for Online Mendelian Inheritance in Man (<https://www.omim.org/>) was developed for the World Wide Web by NCBI (Ada Hamosh). It is a publicly accessible database housing information human genes and genetic phenotypes (Ada Hamosh). OMIM focuses on the connection between phenotype and genotype. It is updated daily and has numerous links to other genetics sites (Ada Hamosh). Currently this database contains a total of 28,187 entries which can be accessed via the OMIM statistics which is the summation of genetic and phenotypic description. Additionally, within an OMIM entry are various other resources or subheadings like Clinical Synopsis and Synopsis of the Human Gene Map (Ada Hamosh).

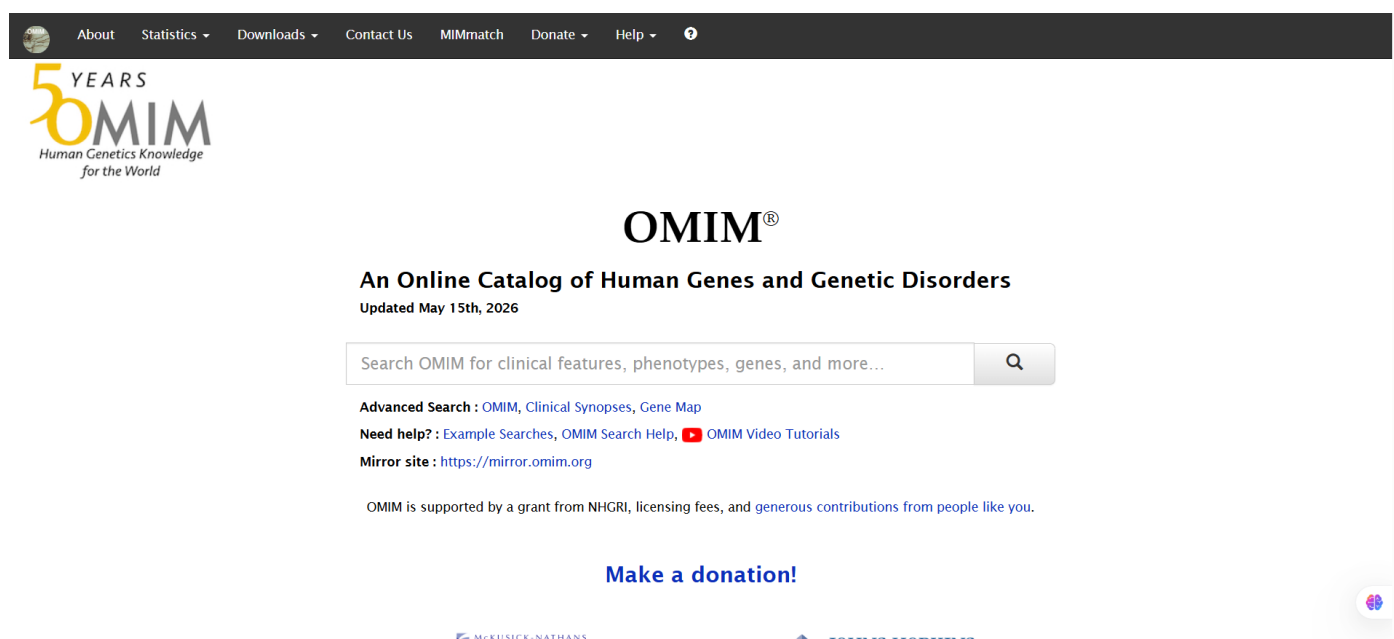


Figure 5: Homepage of OMIM

3.1.6 PubChem:

PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) is a freely available chemical database which houses information on a plethora of chemical compounds and how they react with the body (Sunghwan Kim).

This database curate's information related to a chemical from over a 750 data sources (Sunghwan Kim). One may obtain descriptive datasets and millions of complex structures at no cost. PubChem has tiny compounds with less than 1,000 bonds and 100 atoms, as well as several descriptions of substances. There are three dynamic databases within PubChem i.e Compounds, Substance and BioAssay. Substance database is publicly accessible repository which houses information on chemicals deposited by the user having about 293 million entries (Sunghwan Kim). Compound database is a slightly improved version of the Substance database as it contains only the unique, pure and characterized chemical structures extracted from Substance database, it has about 111 million entries (Sunghwan Kim). BioAssay database as the name implies stores information related to biological assays and tests. This database has about 1.25 million entries (Sunghwan Kim). With support for SMILES/SMARTS and InChI, PubChem has its own online molecule editor that lets users search for structures and fragments by importing and exporting files in all popular chemical file formats (Sunghwan Kim).

3.1.7 CTD:

CTD (<https://ctdbase.org/>) is a massive, open-access database dedicated to better understanding how environmental exposures influence human health (C. J. Allan Peter Davis). It provides information on chemical-gene/protein interactions (C. J. Allan Peter Davis), chemical-disease (C. J. Allan Peter Davis), and gene-disease linkages (C. J. Allan Peter Davis), which has been manually curated (C. J. Allan Peter Davis). These data are combined with functional and pathway data to help researchers build theories about the mechanisms causing diseases that are influenced by the environment (C. J. Allan Peter Davis).

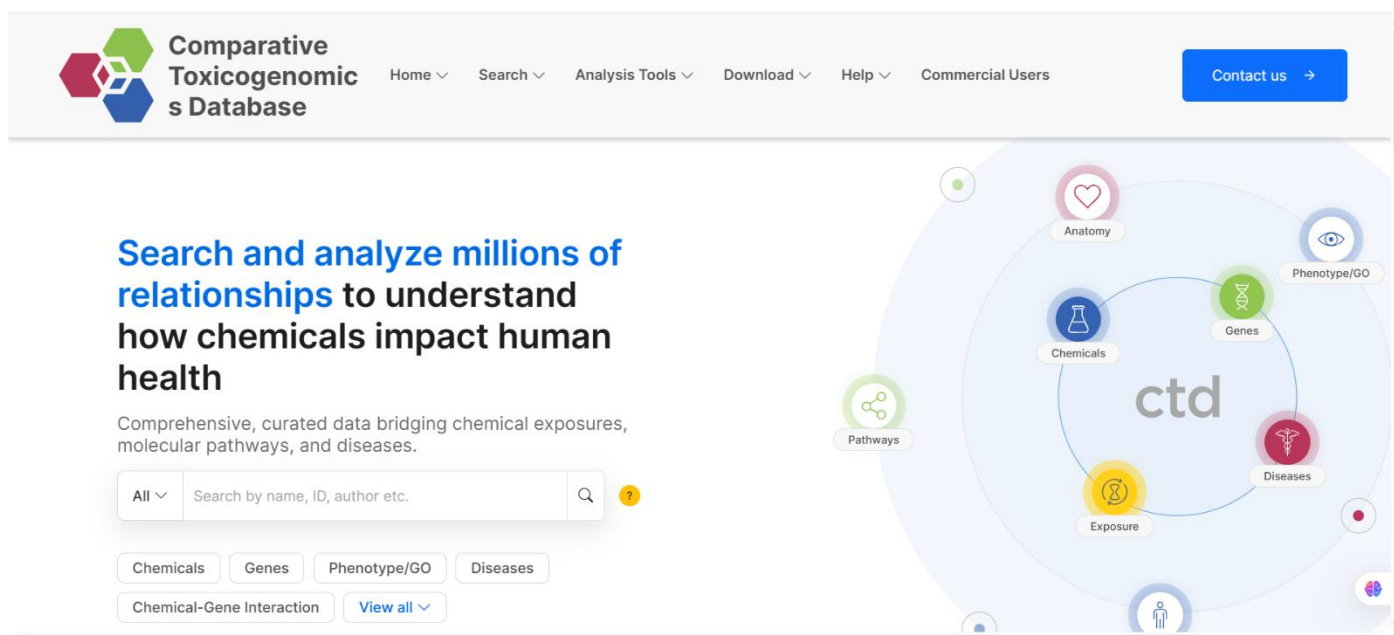


Figure 7: Homepage of Comparative Toxicogenomics Database (CTD)

3.1.8 STP:

STP is the abbreviation for Swiss Target Prediction (<http://www.swisstargetprediction.ch/>) a web-based tool for predicting ligand-based targets for bioactive small molecules (Antoine Daina).

Figure 8: Homepage of Swiss Target Prediction Database

3.1.9 STITCH:

STITCH (<https://stitch-db.org/>) short for Search Tool for Interacting Chemicals is a database built exclusively for assessing and identifying all the interactions occurring between proteins and small molecules. STITCH is the amalgamation of a myriad sources (Vivian Law). STITCH (Damian Szklarczyk).

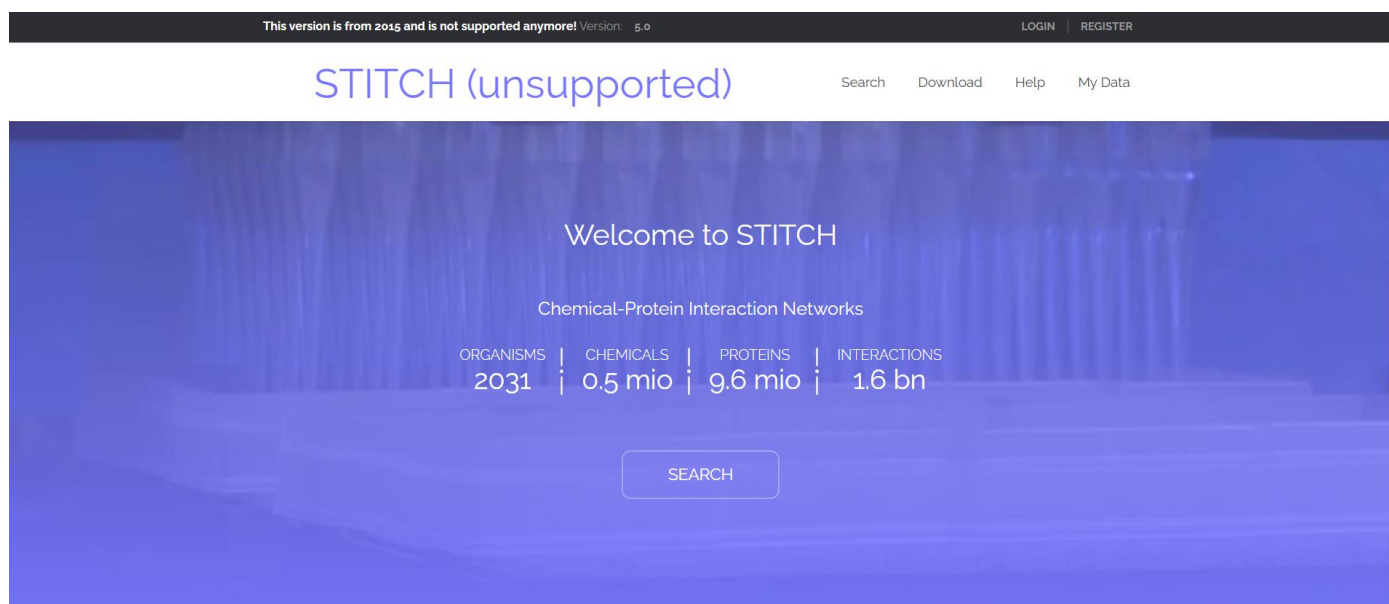


Figure 9: Homepage of STITCH

3.1.10 UniProtKB:

UniProt is a database that houses sequences (T. U. Consortium.) and functional information of proteins (F. G. Rasko Leinonen) (Shadab Ahmad) (T. U. Consortium.) (<https://www.uniprot.org/>). Over 100 people are active in various responsibilities at the three institutes, including database curation, software development, and support (Shadab Ahmad). As of now there are a total of 574,627

total entries in Swiss-Prot and a total of 202,556,313 total entries in TrEMBL (T. U. Consortium.) (Shadab Ahmad). The Uniprot database is updated once in every eight weeks (Shadab Ahmad).

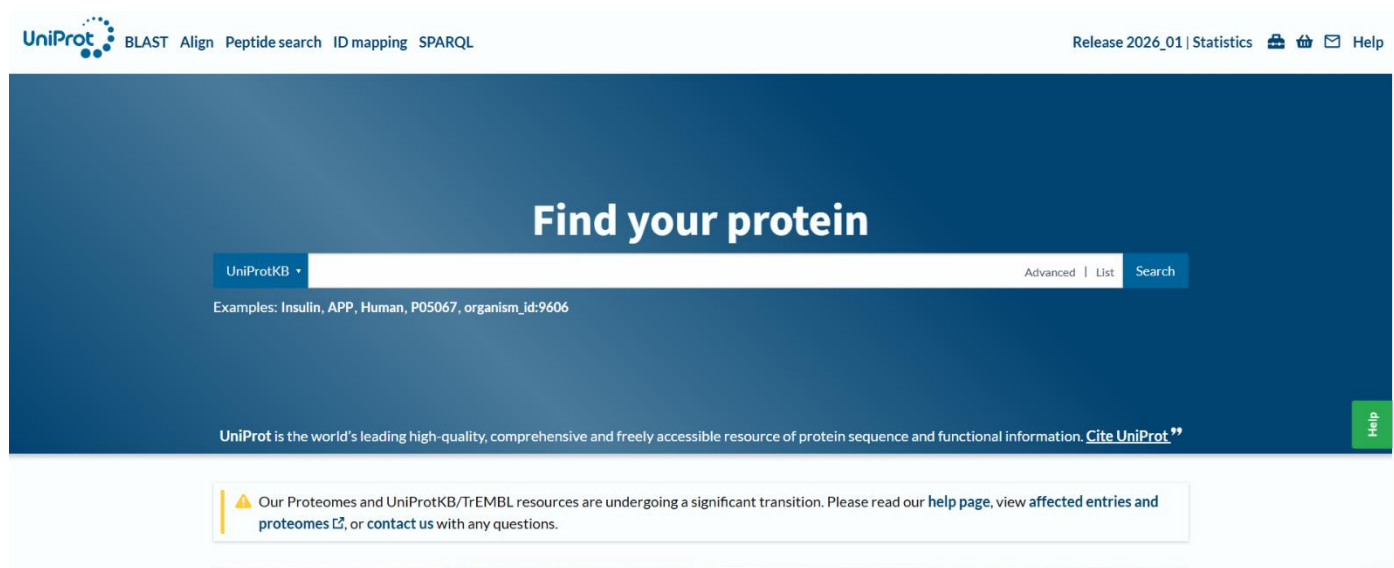


Figure 10: Homepage of UniProt

3.1.11 RCSB PDB:

The Protein Data Bank (PDB) (<https://www.rcsb.org/>) was the first open access repository of digital data in the history of biology and medicine (Helen Berman) (Helen M. Berman). It is now a prominent worldwide source for experimental data essential to advancements in science (Helen M. Berman). A wealth of 3D structure data has been deposited in the PDB, which has helped considerably to increase our understanding of protein architecture (Helen M. Berman). All this has led to recent advances in the prediction of protein structure, greatly accelerated by deep learning and artificial intelligence (Helen M. Berman). The latest version of PDB is currently PDB 3.1 that contains information on 3D structure of 217,157 proteins. And information about 1,068,577 Computed Structure Models (CSM)

(Helen M. Berman). Usually, the data comes from X-ray crystallography, NMR spectroscopy and cryo-electron microscopy (Helen M. Berman).

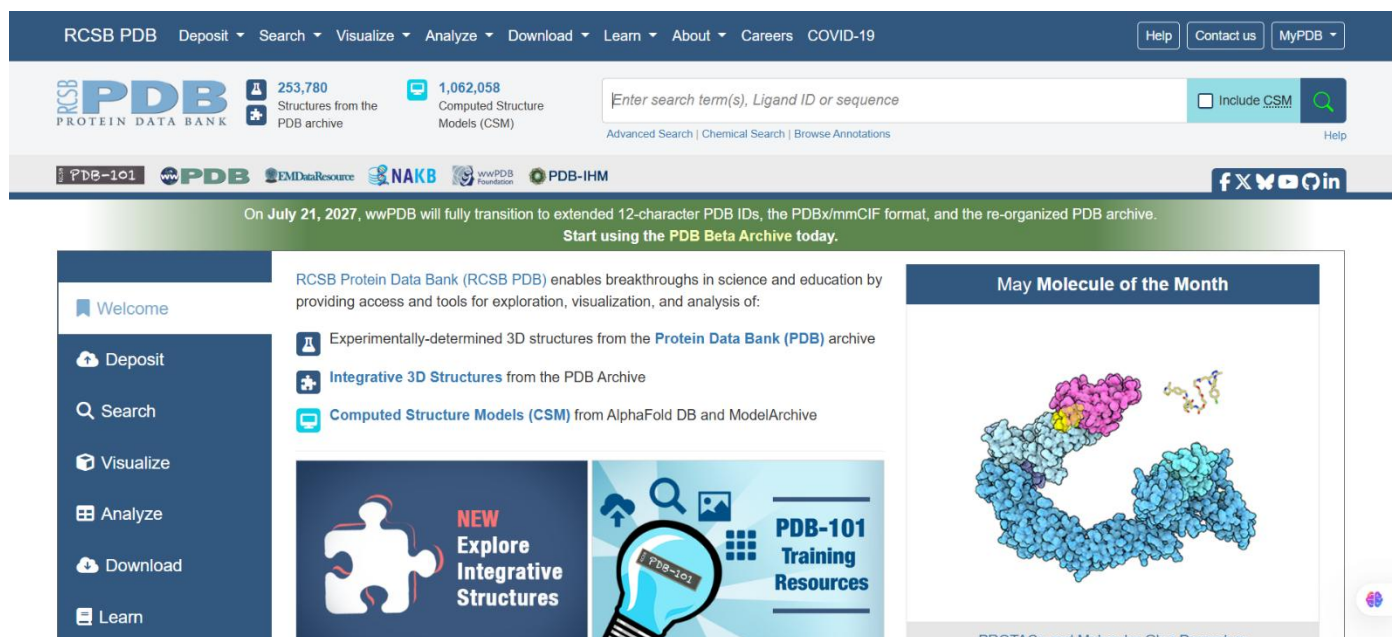


Figure 11: Homepage of RCSB PDB

3.1.12 Discovery Studio Visualizer:

A fully fledged software package, Discovery Studio is tailored to the analysis and modelling of sequences, molecular structures and data of relevance to the life science field (Biovia.). This software provides features such as simple data analysis, data viewing and editing (Biovia.). Data produced by other Discovery data products can be viewed using the free Discovery Studio Visualizer (Biovia.). All other options pale in comparison to tool panels for accessing the tools (Biovia.). Tool panels are used to organize tools into groups for specific tasks. Besides, the tool panels are grouped in large sets, such as receptor-ligand interactions and pharmacophores (Biovia.).

3.1.13 Rstudio:

Rstudio is an Integrated Development Environment (IDE) that is cross-platform and designed to enhance the functionality and usability of the R programming language (Racine.). RStudio offers a unified and easy to use solution, by bringing together key components in a single interface (Racine.). It consists of four large panels: editor for writing scripts, console for running commands,

workspace/history window, and file management, charting, handling packages, and help documentation (Racine.). The design is flexible and can be used in any way to suit individual user habits (Racine.). RStudio is well known for its platform independence, as it supports Windows, macOS and Linux (Racine.). It features a tabbed source editor, syntax highlighting, auto-completion, and even code execution, a searchable command history for easy retrieval and reuse of past code, and built-in code transformation tools (Racine.).

3.1.14 AutoDock Vina:

Molecular docking was underdone using Vina to estimate the binding relationship between the receptor and ligand (O. T. Olson.). Vina predicts the shape of the complex generated and the binding affinity, or in simpler terms, the strength of the interactions that occur between the receptor and the ligand (O. T. Olson.). Used far and wide in drug discovery process (O. T. Olson.). AutoDock Vina is the molecular docking software used which is better than it's predecessor i.e AutoDock 4 in terms of prediction speed and accuracy (O. T. Olson.). Vina utilizes a gradient based optimization algorithm to asses the binding conformations accurately. Vina uses a scoring function which evaluates the receptor ligand interaction based upon hydrogen bonding, hydrophobic interactions, steric hindrance and ligand flexibility (O. T. Olson.).

3.1.15 GTex:

Genotype-Tissue Expression (GTEx) database (<https://gtexportal.org/home/>) helps in understanding how different gene expression is

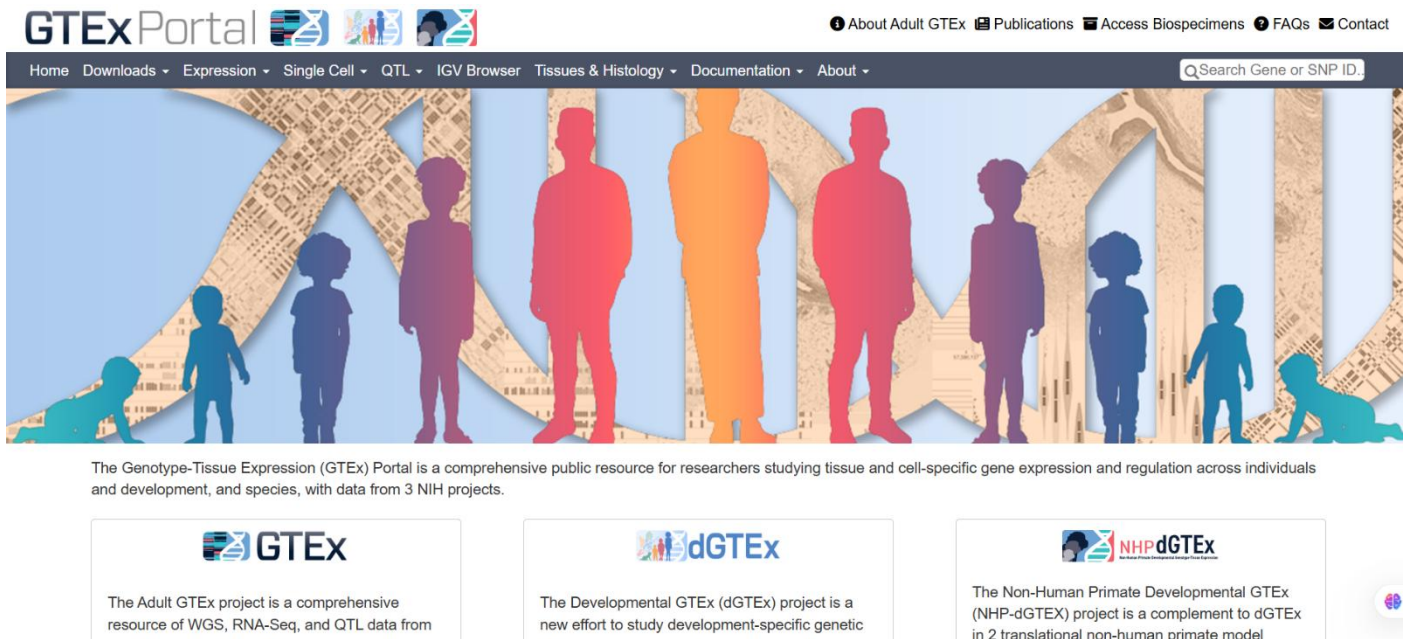


Figure 12: Homepage of GTEx

across males and females established (A. S. Wayez Naqvi) (Moore.).

3.2 Methodology:

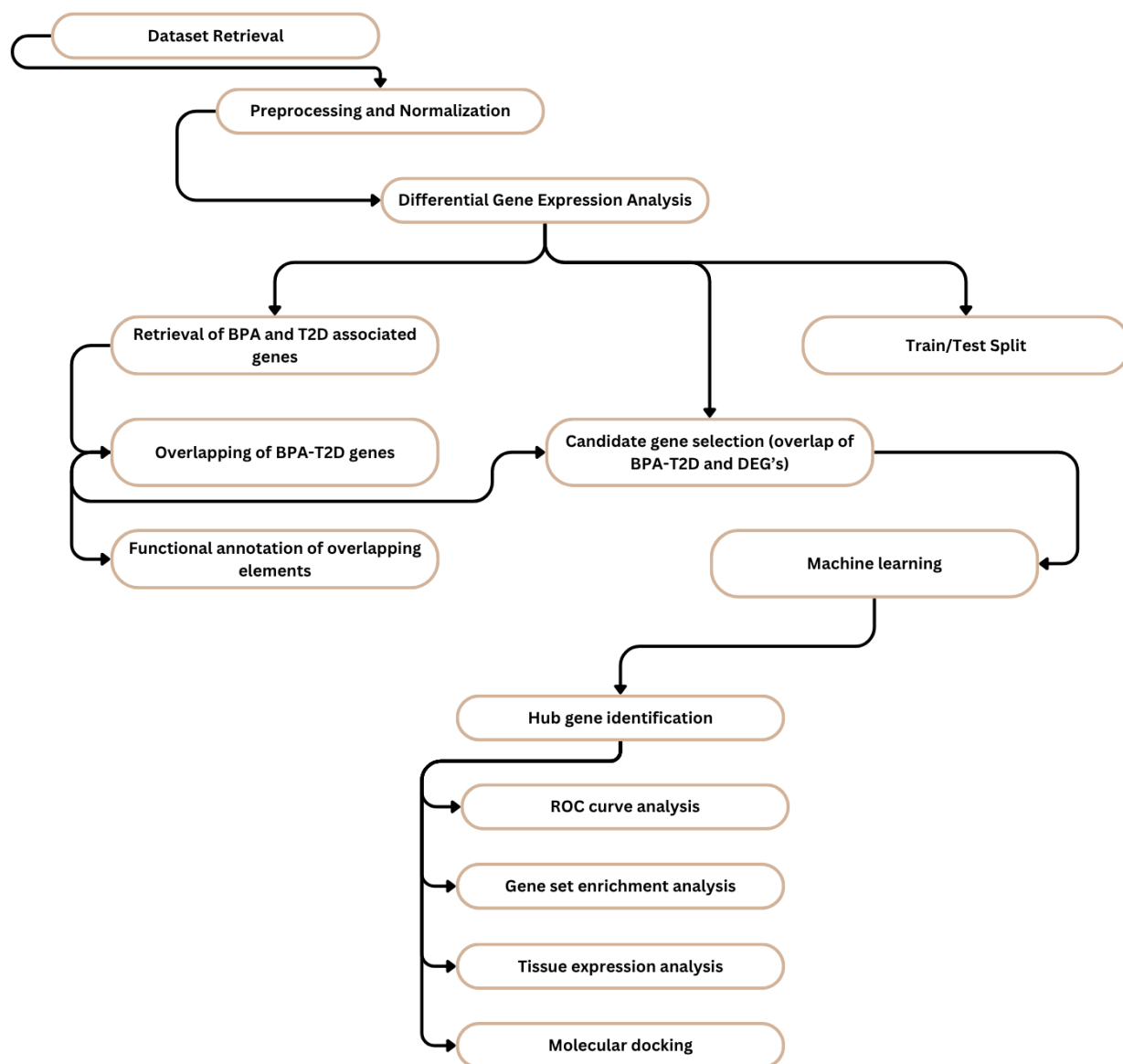


Figure 13: Flowchart depicting the pipeline followed in this study (P. G. Wayez Naqvi)

3.2.1 Dataset Retrieval

Through GEO (Tanya Barrett) and SRA (H. S. Rasko Leinonen) of NCBI (Conrad L Schoch) datasets were retrieve datasets for this study. Selection was based on the criteria that the studies had type 2 diabetes as the disease condition, had blood as the sample source and had transcriptomic data type

3.2.2 Data Preprocessing:

Following dataset retrieval, the raw counts file were imported and mapped against org.Hs.eg.db to convert entrez id to gene symbols. Following of which only the common genes across all datasets were identified and retained in the merged expression matrix. Then the low count genes were removed to reduce background noise. Using DESeq2 package normalization and variance stabilizing transformation was done followed by which using sva package batch correction was done (Yuqing Zhang).

3.2.3: Differential Gene Expression Analysis:

Following preprocessing, normalization and batch correction, differential gene expression analysis was underdone. Log2 fold change shrinkage was done using apeglm's lfcshrink function and genes with null p-values were filtered out.

3.2.4 Retrieval of BPA targets and T2DM genes:

Next, the gene targets associated with BPA were retrieved using Swiss Target Prediction database, Comparative Toxicogenomic Database and STITCH. While using GeneCards, Malacards and OMIM, genes related to Type 2 Diabetes Mellitus were retrieved. The obtained lists were then overlapped to identify the intersecting elements between BPA and T2DM genes.

3.2.5 Functional annotation:

Next, functional enrichment analysis was carried out on the BPA-T2D intersecting genes (Bobak.). GO and KEGG enrichment (Bobak.) was underdone using clusterProfiler package (Bobak.) (Weiqing Yang).

3.2.6 Identification of candidate genes:

The genes obtained from differential gene expression analysis were then overlapped with intersecting elements of BPA and T2D. The interacting entries were then retained and utilized for further analysis.

3.2.7 Training and Validation Set Construction:

Following DEG analysis, samples were distributed into training and test set (Weiqing Yang) through random stratified sampling (Weiqing Yang) at a ratio of 7:3 (Weiqing Yang).

3.2.8 Machine learning analysis:

RF (Breiman.), SVM-RFE (Mei-Ling Huang) and LASSO (Monica M. Vasquez) were used employed for identifying candidates genes. (Weiqing Yang).

3.2.9 Diagnostic validation using ROC curve:

Using pROC package (Linlin Song), the hub genes obtained after machine learning approaches (P. G. Wayez Naqvi) evaluated on the test set to assess the diagnostic performance (Linlin Song).

3.2.10 Pathway Analysis:

Using the clusterProfiler package, GSEA was underdone (Weiqing Yang) to map out the associated pathways (Weiqing Yang). Genes were ranked using Wald statistic first and then pathways were evaluated using adjusted p-value and normalized enrichment scores. (Weiqing Yang)

3.2.11 Tissue expression analysis:

Using GTEx database, expression of the potential diagnostic biomarkers across different human tissues was assessed.

3.2.12 Molecular Interaction Analysis:

Molecular docking was performed to assess the binding interaction occurring between bisphenol A and potential diagnostic biomarkers using AutoDock Vina.

RESULT & DISCUSSION

4.1 Dataset Retrieval:

Through GEO and SRA of NCBI datasets were retrieved for this study. Table 1 shows the summary of the four datasets utilized in this research. Selection was based on the criteria that the studies had type 2 diabetes as the disease condition, had blood as the sample source, and had transcriptomic data type. Three out of four datasets were manually filtered to remove samples associated with some other disease condition apart from Type 2 diabetes.

Table 1: Summary of the selected datasets available on GEO with their respective Bioproject and SRA accession ID's, and sample sizes.

GEO Accession ID	Bioproject Accession ID	SRA Accession ID	Sample Size	Sample Size (after filtering)	Total Patient Samples	Total Control Samples	Reference
GSE154881	PRJNA647610	SRP273003	15	10	5	5	(Aleksandr A Tkachenko)
GSE153315	PRJNA642130	SRP268985	30	30	20	10	(Aleksandr A Tkachenko)
GSE184050	PRJNA650123	SRP275718	116	116	50	66	(Hung- Hsin Chen)
GSE114192	PRJNA470512	SRP144997	249	88	52	36	(Clare Eckold)

4.2 Data Preprocessing:

Following dataset retrieval, the raw counts file imported and mapped against org.Hs.eg.db to convert entrez id to gene symbols. Following of which common genes across all datasets were identified and retained in the merged expression matrix. Then the low count genes were removed to reduce background noise. Using the DESeq2 package, normalization and variance stabilizing transformation was done and using sva package batch correction was done (Yuqing Zhang). Figures 14 and 15 depict the PCA's before and after batch correction respectively.

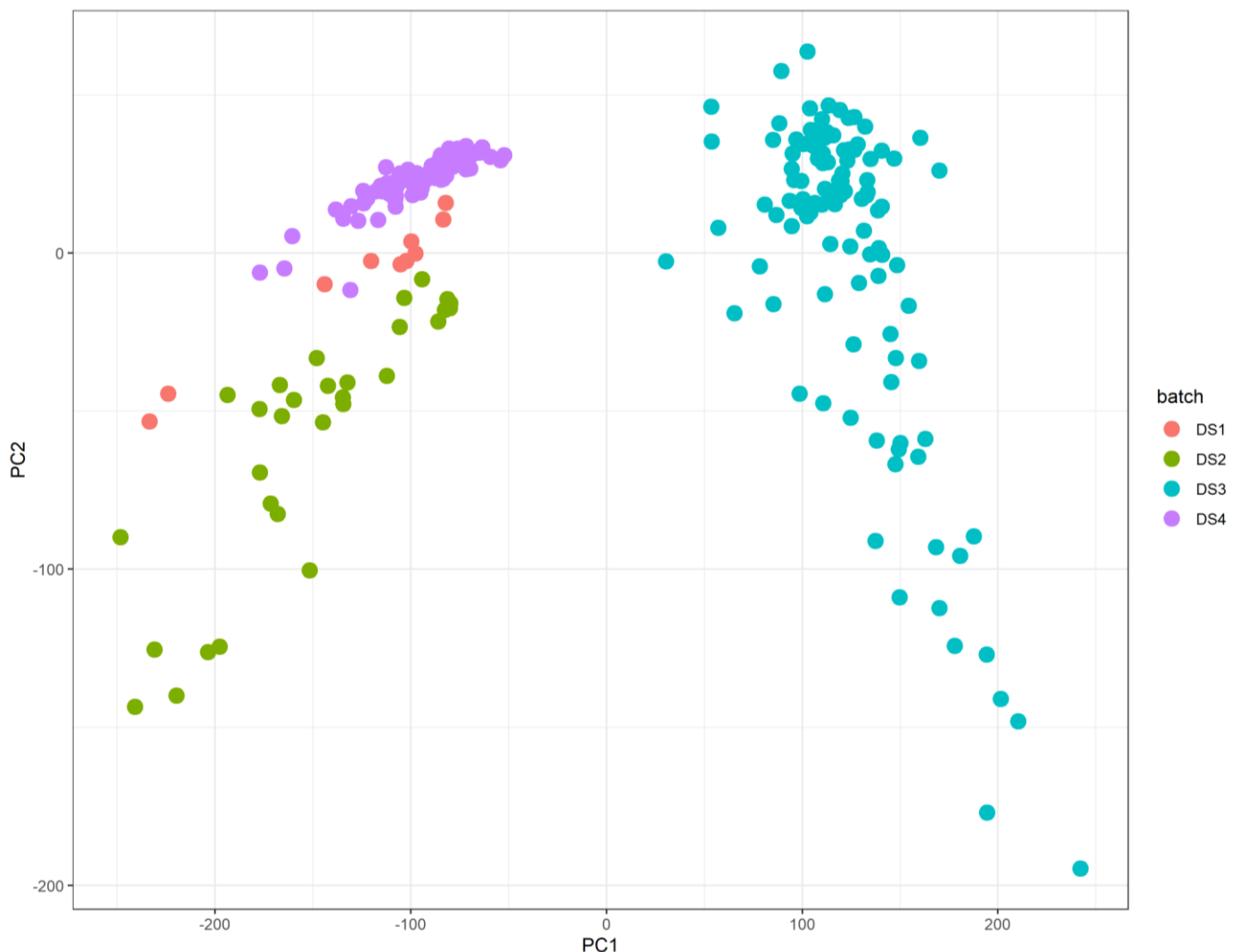


Figure 14: Initial PCA plot of the T2D datasets. Here DS-1 represents samples of GSE154881, DS-2 represents samples of GSE153315, DS-3 represents samples of GSE184050, and DS-4 represents samples of GSE114192

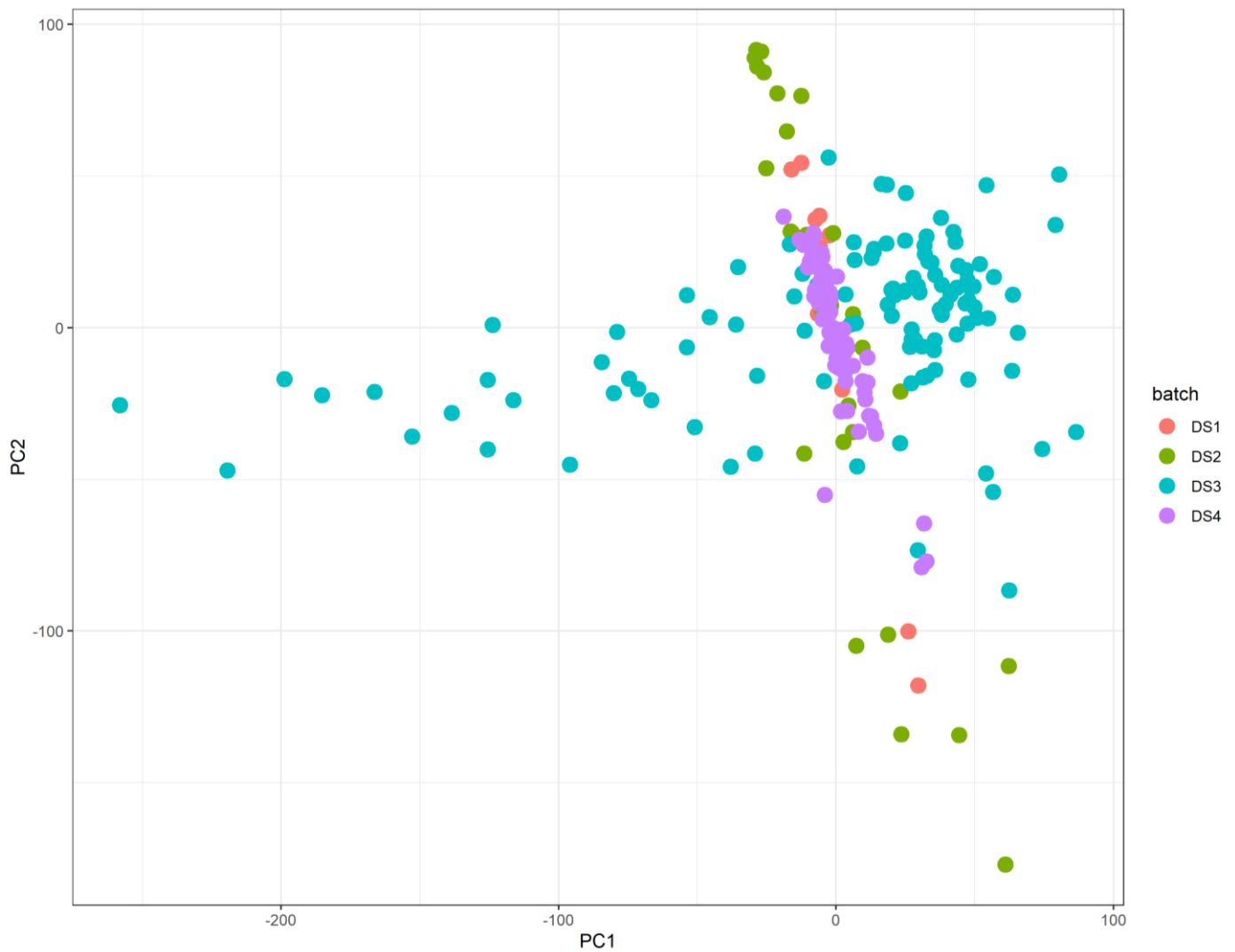


Figure 15: PCA plot after batch correction. Here DS-1 represents samples of GSE154881, DS-2 represents samples of GSE153315, DS-3 represents samples of GSE184050, and DS-4 represents samples of GSE114192

4.3 DEG Analysis:

Following data processing, DEG analysis was done. The `apeglm`'s `lfcshrink` function was used to do log2-fold change shrinking. By setting the p-value < 0.05 (P. G. Wayez Naqvi) and log 2-fold change value > 0.25 (P. G. Wayez Naqvi) 355 DEG's were identified (P. G. Wayez Naqvi) (figure 16 show the results visualized in a volcano plot) (P. G. Wayez Naqvi).

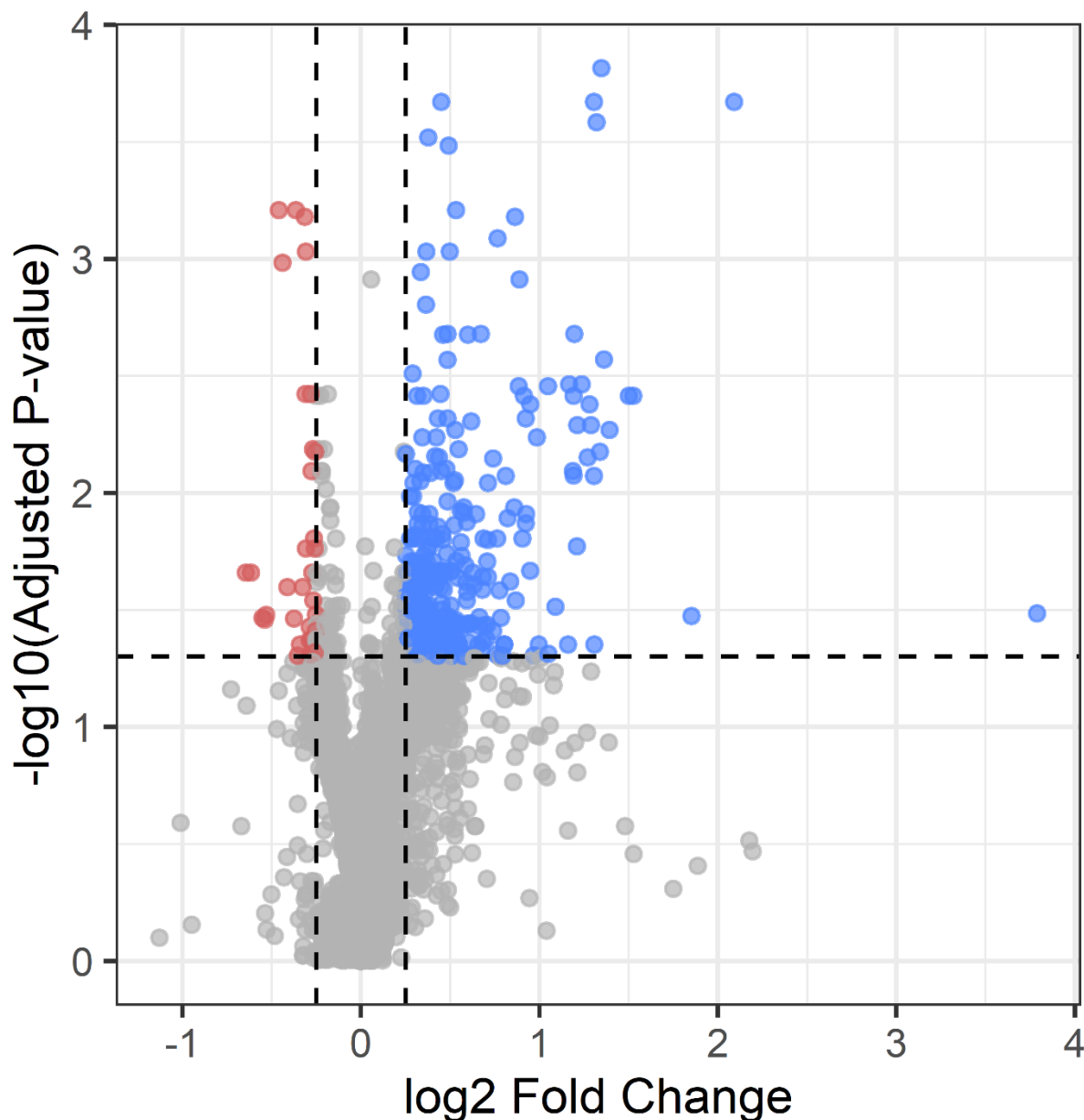


Figure 16: The volcano graphic depicts the differentially expressed genes retrieved from DESeq2 (P. G. Wayez Naqvi)

4.4 Retrieval of BPA targets and T2DM genes:

For retrieving the gene targets associated with BPA, three databases were utilized i.e Swiss Target Prediction (STP), Comparative Toxicogenomic Database (CTD) and STITCH (Weiqing Yang). The obtained lists from were then merged and de-duplicated to create a non-redundant list (Chenglin Wang). Similarly, using GeneCards, Malacards and OMIM, Type 2 Diabetes genes were retrieved, out of which only the protein coding genes were taken into consideration (Chenglin Wang). The obtained list was then filtered as well to create a non-redundant list (Bo Wu). Finally, the 19063 BPA genes and

4148 T2D genes were compared against each other with the intersection yielding 3885 genes (as shown in figure 17) These shared genes were then utilized for further research.

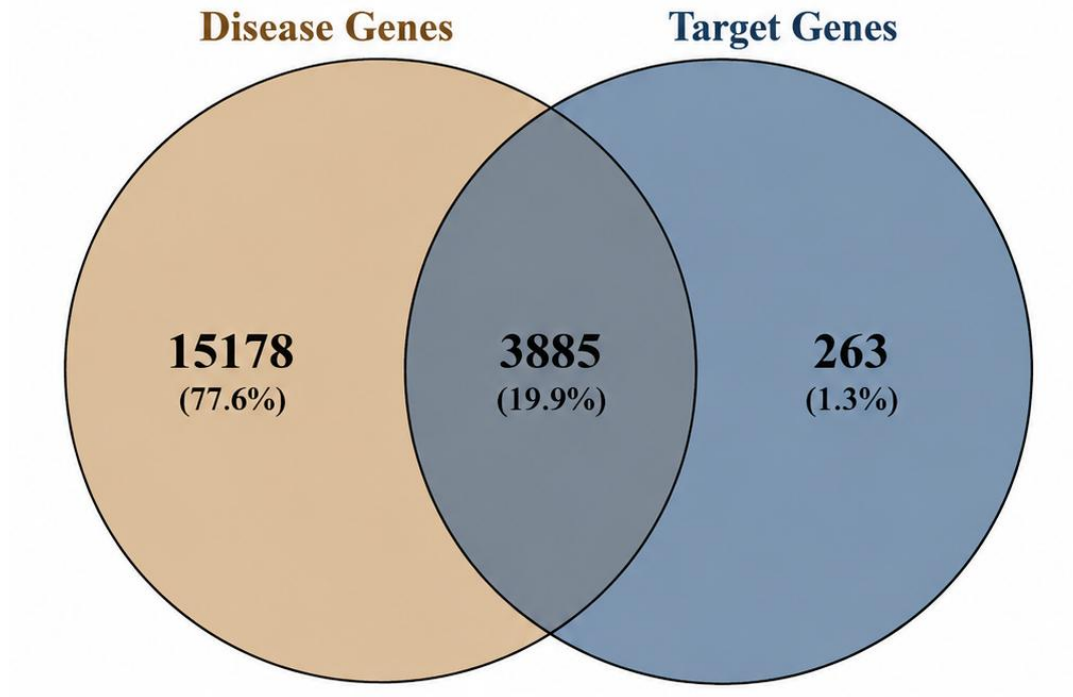


Figure 17: Venn diagram showing the overlapping genes between BPA target genes and T2DM disease genes

4.5 Functional annotation of overlapping elements:

Next, the overlapping elements of BPA-T2D functionally annotated using the clusterProfiler package (Linlin Song) [using gene ontology and KEGG terms] (figures 18 and 19 depict the bar plots for GO and KEGG enrichment analysis). Biological processes like metabolic regulation, hormone responsiveness, oxidative stress, inflammation, and xenobiotic response were significantly enriched. KEGG pathway analysis revealed pathways like PI3K-Akt signaling, MAPK signaling, AGE-RAGE signaling, cAMP signaling, and apoptosis to be significantly associated. Membrane-associated structures and vesicle-mediated transport processes were the cellular components associated with BPA-T2D intersecting genes while the molecular functions cytokine signaling, growth factor interactions, and transcriptional regulation were significantly enriched.

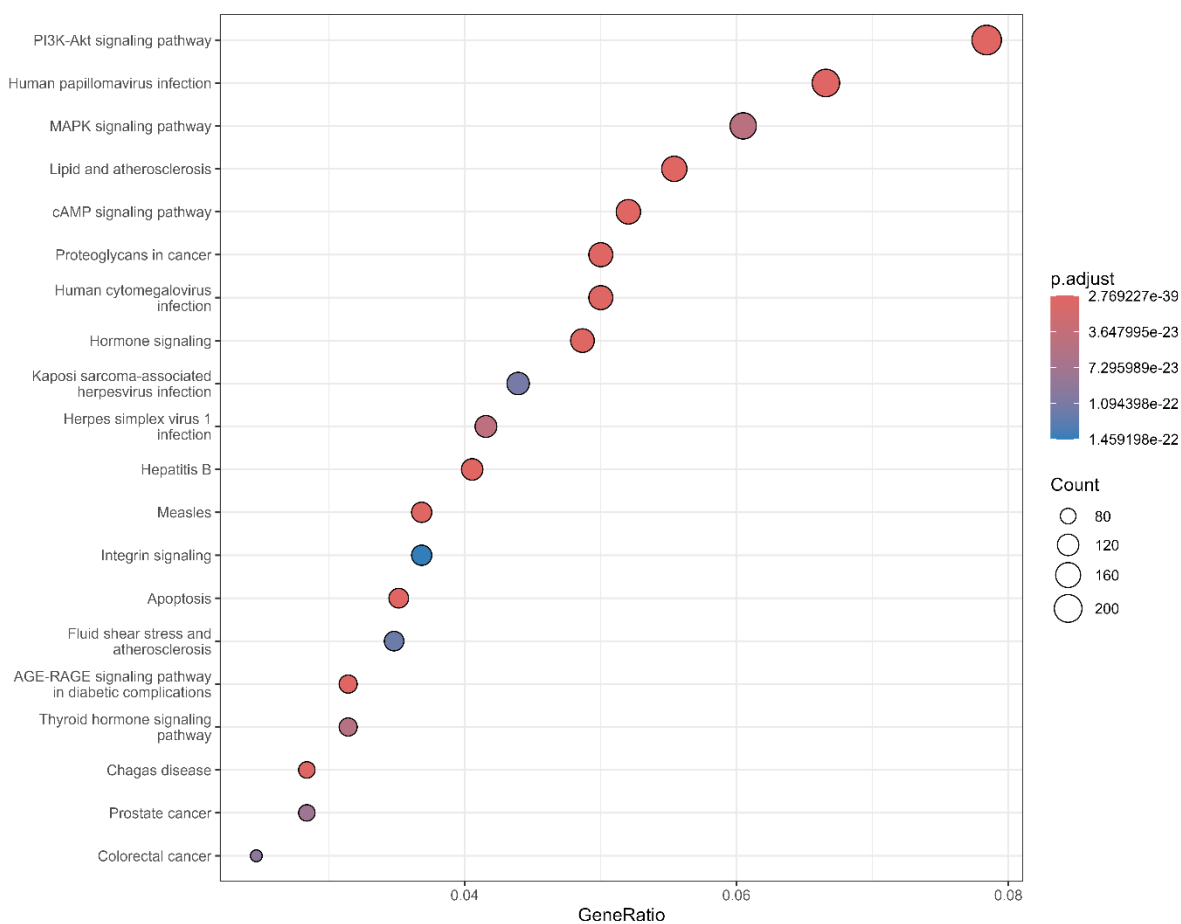


Figure 18: Functional profiling of the intersecting BPA T2D genes using KEGG terms

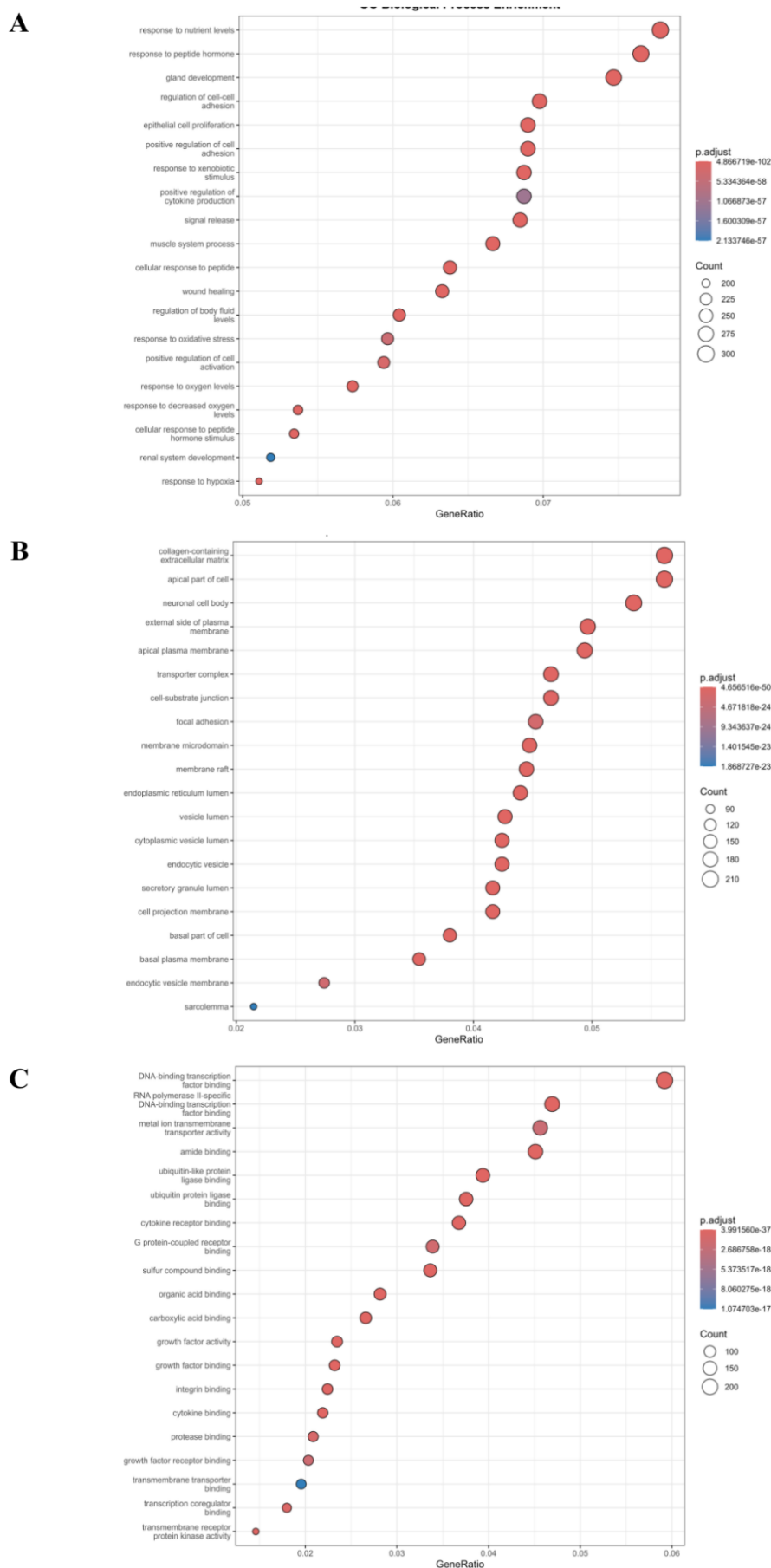


Figure 19: Functional profiling of the intersecting BPA T2D genes using Gene ontology terms

4.6 Identification of candidate genes:

The genes obtained from DEG's were then compared BPA-T2D intersecting genes which resulted in the identification of 52 overlapping genes (shown in table 2) which were then utilized for further analysis.

Table 2: Table showing the 52 genes obtained after the overlap of DEG, BPA and T2D genes.

Gene Symbols			
HNF1A	PAPSS2	OTOF	DRD4
RETN	IGF2R	NEK8	CCDC40
PRKAR1A	AMPD2	NDUFS8	GP1BB
ATP7A	LTA	MMP8	RAPGEF2
CXCL8	FXVD2	LDB3	CERS2
CDKL5	SOCS1	ABCG1	MORC3
COL9A2	MAPK10	S1PR2	GP9
CPT1A	CCL3	LIF	SIRT4
ATP8B1	CRB2	FOXH1	INHBA
LCN2	BAD	ARMC5	RDH5
JAK1	RRAD	EDARADD	H2AC20
HBA2	ANGPTL4	LAMP1	IFI27
FOXC1	CLU	GLUL	PXDNL

4.7 Training and Validation Set Construction:

Following preprocessing and batch correction, samples were distributed into training and test set (Weiqing Yang) through random stratified sampling (Weiqing Yang) at a ratio of 7:3 (Weiqing Yang).

4.8 Machine learning based hub gene identification:

RF (Breiman.), SVM-RFE (Mei-Ling Huang) and LASSO (Monica M. Vasquez) were employed for identifying candidates genes. (Weiqing Yang), feature selection and model building. In RF, the randomForest package was used with 1000 trees to rank genes depending on the mean decrease scores (figure 20 shows the feature importance curve with top 10 genes). A total of 20 genes were obtained through this. Next, the out of bag error estimation was done (shown in figure 21) which yielded a value of 30.59% (Weiqing Yang).

Next, SVM-RFE was used to determine the number of genes or features retained during recursive feature selection and the least informative genes were removed. As shown in figure 22, an accuracy of 69.9% was obtained when 9 genes were taken into consideration (Weiqing Yang).

Lastly, LASSO regression model was built with a 10-fold cross validation using glmnet package of R. Lambda.min was used for assessing the penalty strength, out of the 52 candidate genes 18 were retained as they retained non-zero coefficients while the remaining genes shrank towards zero coefficients (shown in figure 23 and 24). Outcomes of these three approaches (shown in table 3) were then overlapped using a Venn diagram (shown in figure 25) to identify the genes uniformly distributed which were identified to be ABCG1, AMPD2 and HBA2 (Weiqing Yang).

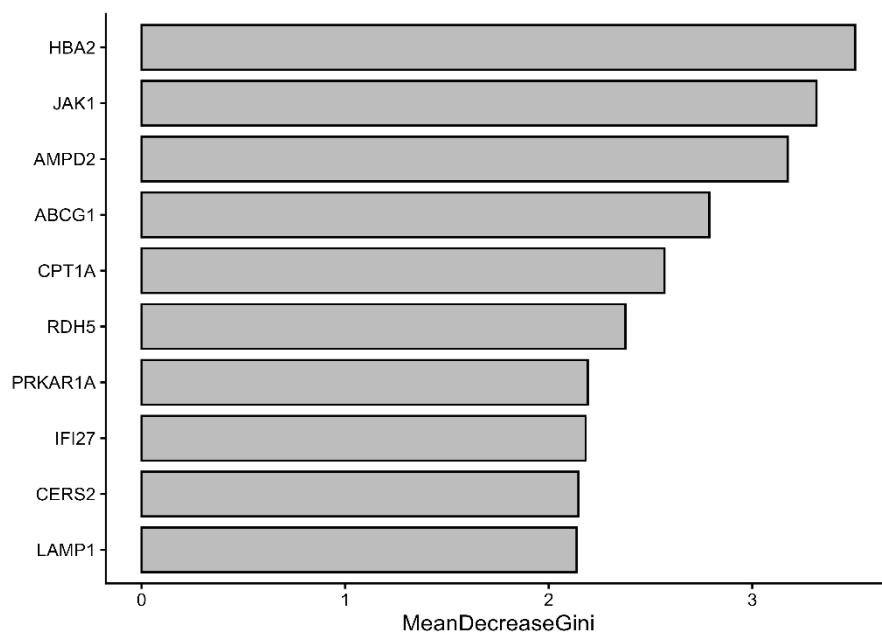


Figure 20: Feature importance curve of the top 10 Random forest genes

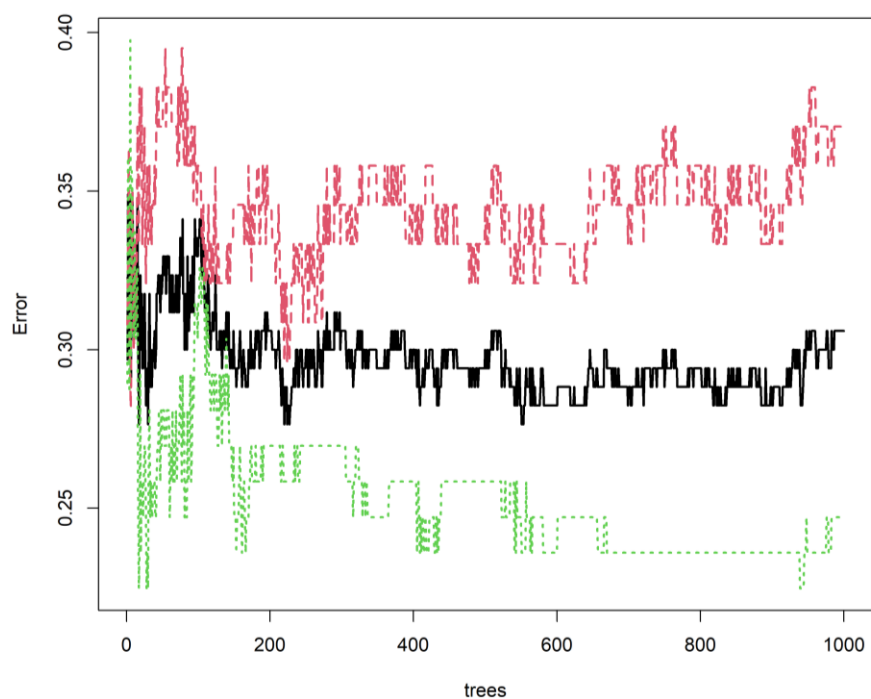


Figure 21: Out of bag error estimation of the Random Forest model

Table 3: Table showing the set genes obtained through the use of machine learning approaches.

RF Genes	SVM Genes	LASSO Genes
HBA2	AMPD2	ABCG1
JAK1	JAK1	AMPD2
AMPD2	HBA2	CCDC40
ABCG1	ABCG1	CDKL5
CPT1A	PRKAR1A	CPT1A
RDH5	LAMP1	DRD4
PRKAR1A	IGF2R	FXYD2
IFI27	CERS2	GLUL
CERS2	CRB2	GP1BB
LAMP1		H2AC20
RAPGEF2		HBA2
CRB2		IFI27
LCN2		MMP8
SOCS1		PXDNL
CCDC40		RDH5
LIF		RETN
GP1BB		RRAD
DRD4		SOCS1
IGF2R		
S1PR2		

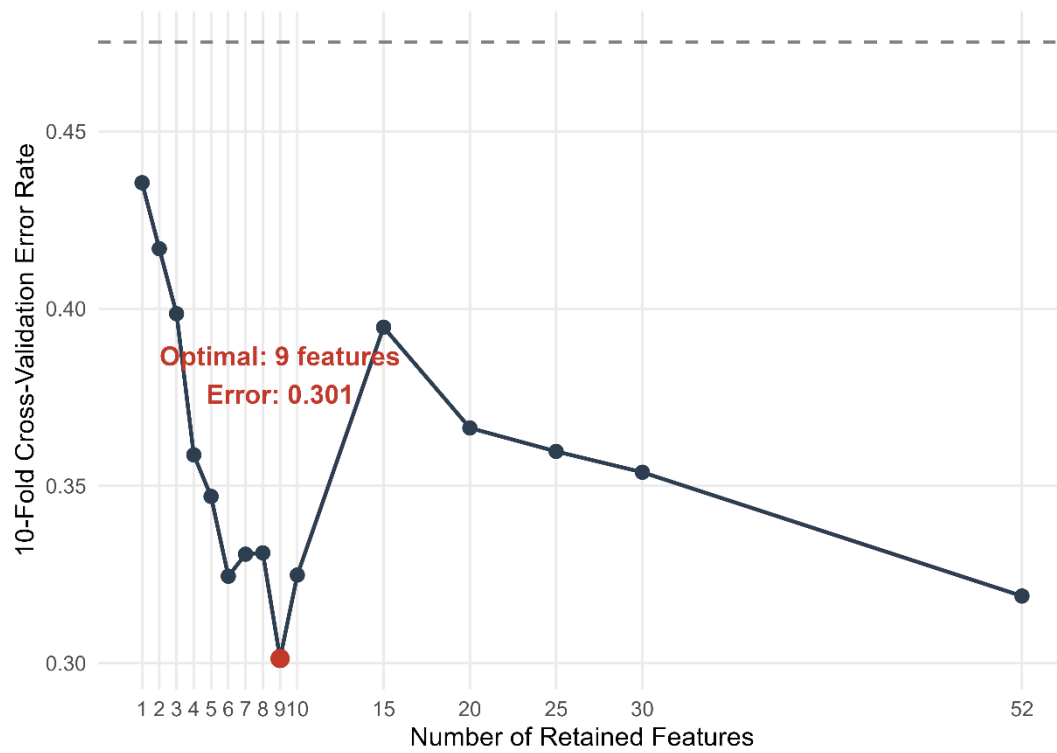


Figure 22: Feature selection performance of the SVM-RFE using 5-fold cross-validation

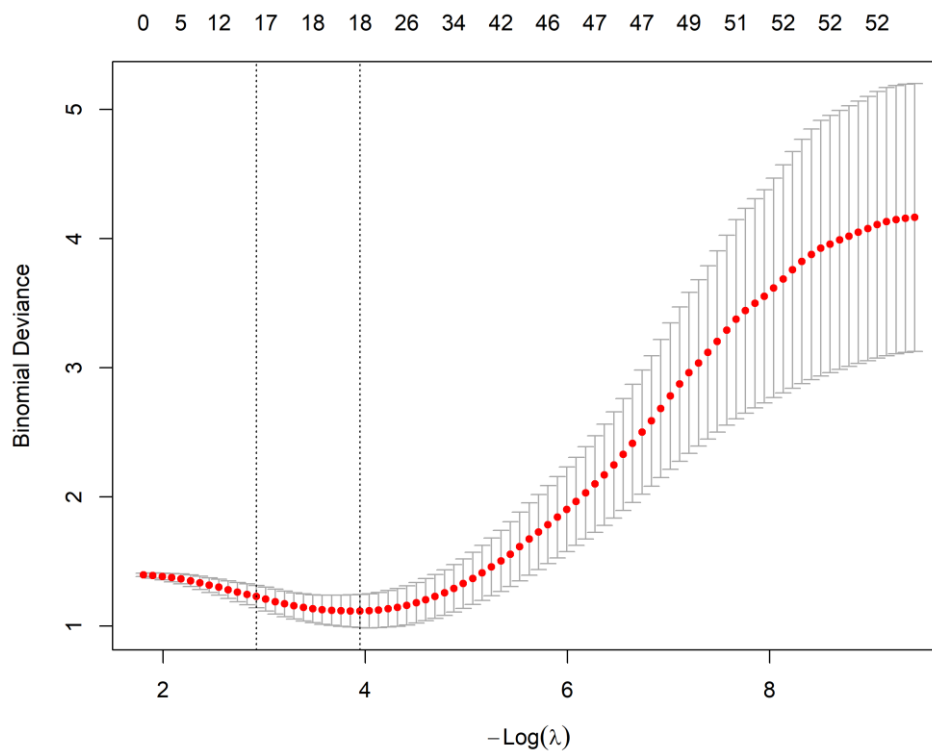


Figure 23: 10-fold cross-validation error curve of LASSO regression analysis

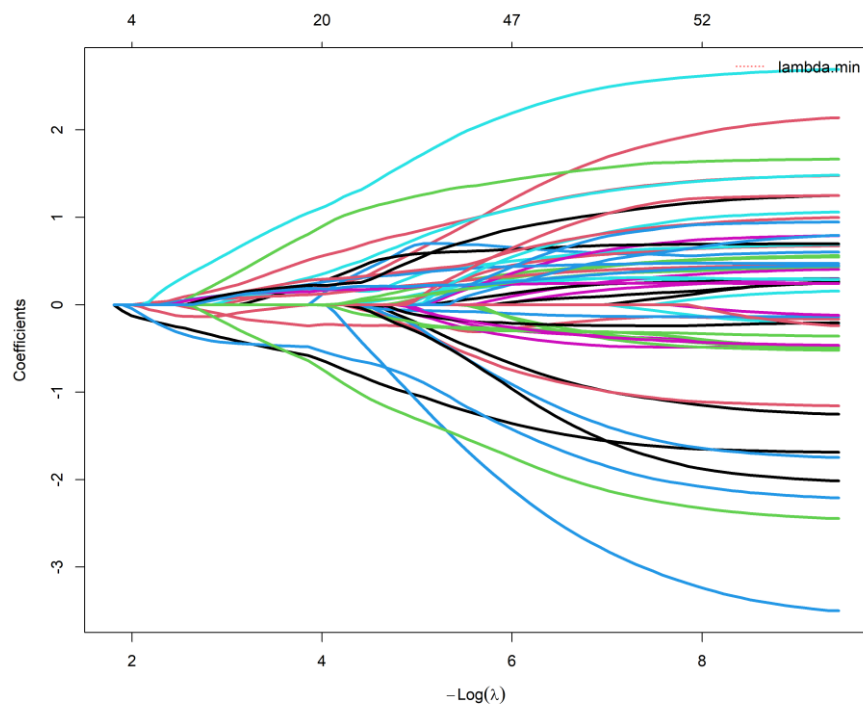


Figure 24: Coefficient trajectory plot of LASSO model, where each coloured lines represents the coefficient of each candidate genes taken in consideration

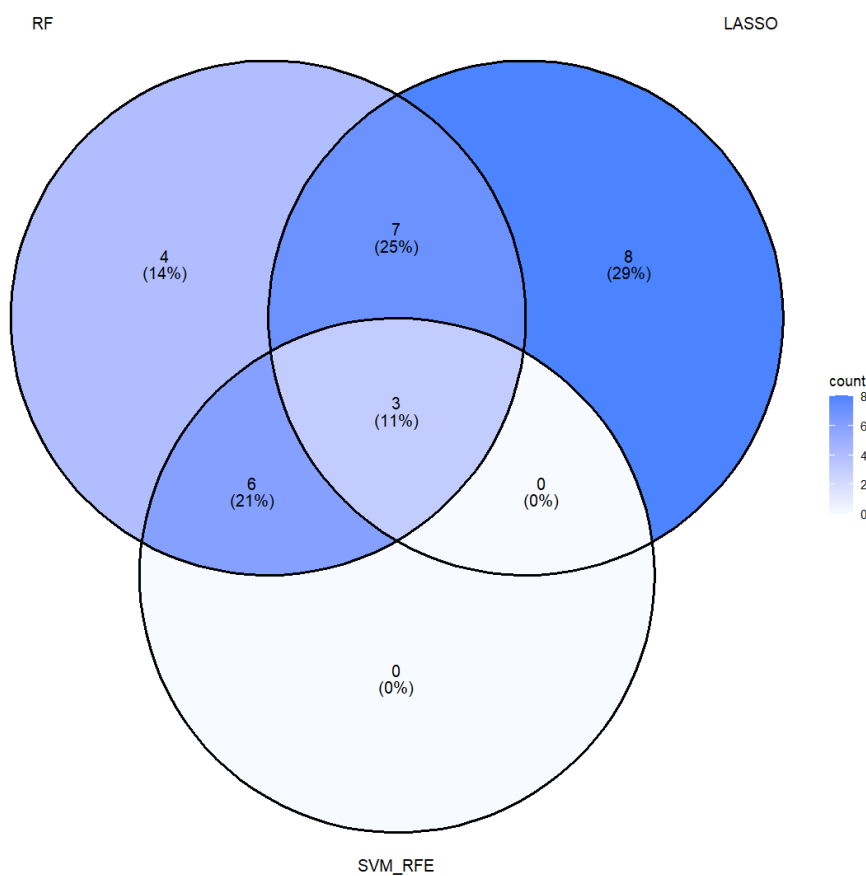


Figure 25: Venn diagram showing overlapping genes obtained from the three machine learning approaches

4.9 ROC curve analysis and diagnostic validation:

Through the use the pROC package, ROC curves for the candidate hub genes were generated using the test set. Wilcoxon rank-sum test was also done (Shaoheng Liang) to rectify the gene expression differences between control and disease groups (Shaoheng Liang). The genes yielded the p-values 5.12×10^{-7} (APMD2), 2.69×10^{-8} (ABCG1) and 2.98×10^{-7} (HBA2) respectively (shown in figure 27) (Yuqin Tang). The genes ABCG1, AMPD2 and HBA2 yielded the following AUC values on the test set respectively: 0.737 (ABCG1), 0.634 (AMPD2) and 0.677 (HBA2) (shown in figure 27). Furthermore, a combined logistic model with all the three hub genes yielded an AUC value of .0745 (shown in figure 26). These findings indicate that AMPD2, ABCG1, and HBA2 possess potential diagnostic value and may contribute to of BPA-associated T2DM.

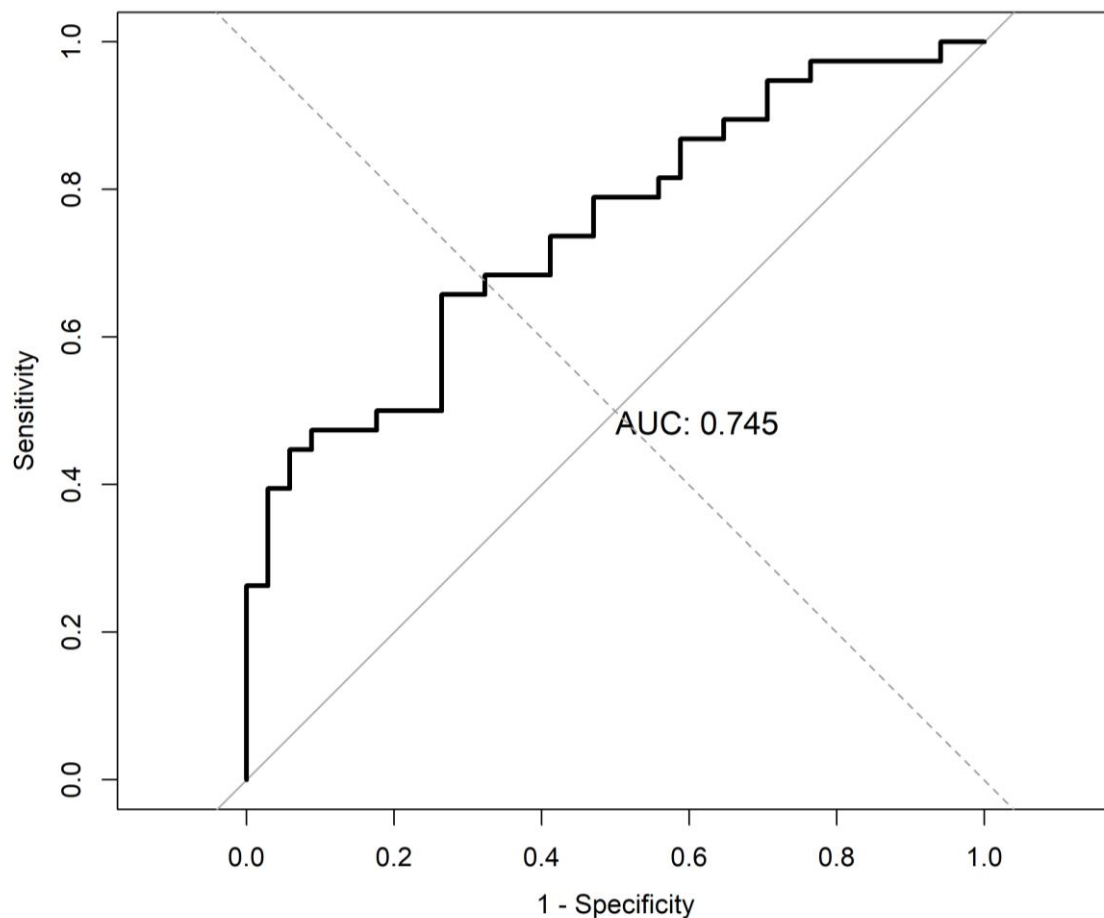


Figure 26: Figure showing the 3 gene ROC curve

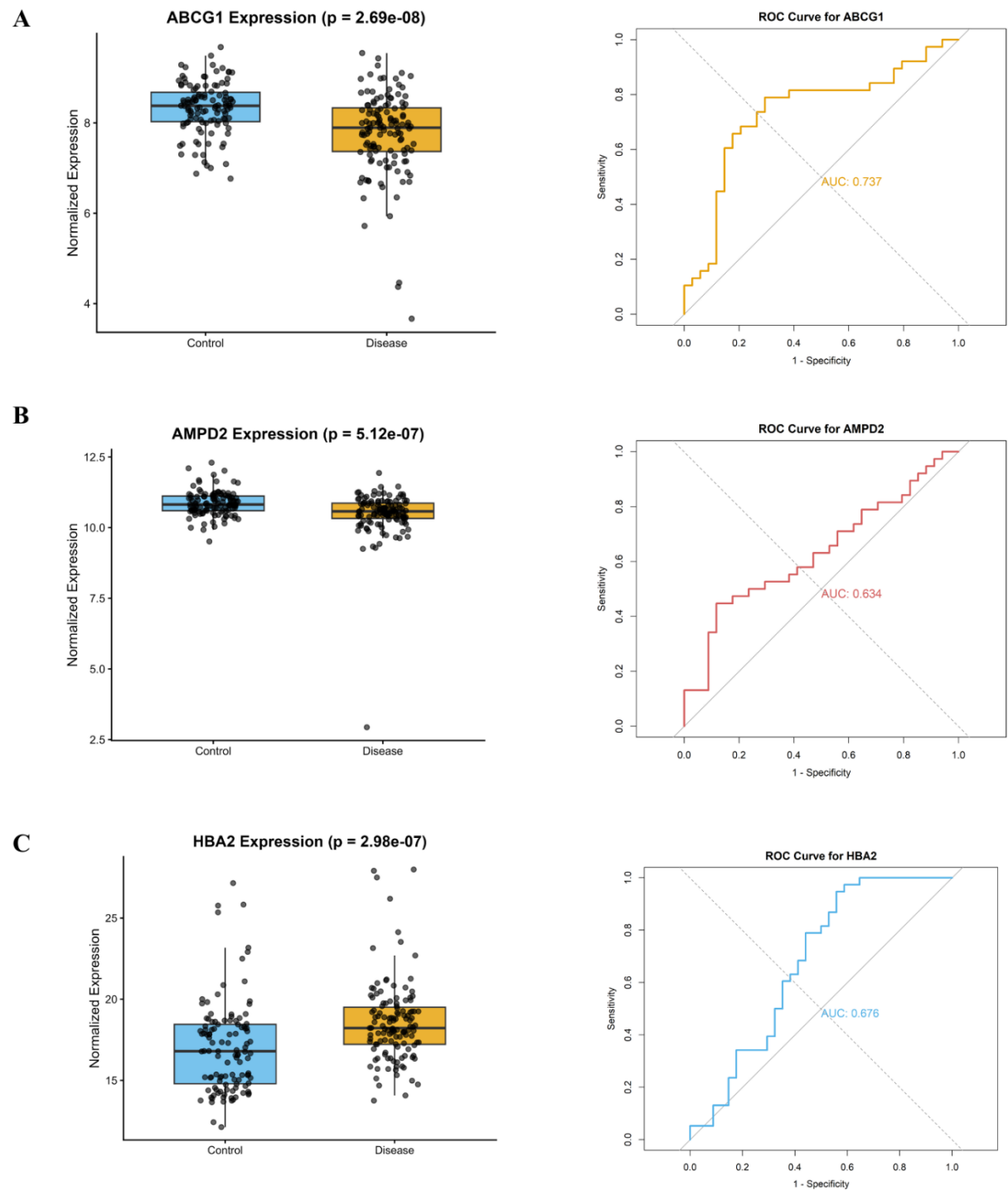


Figure 27: Figure showing the jittered boxplot and single gene ROC curve of (A) ABCG1; (B) AMPD2; and (C) HBA2

4.10 Gene set enrichment analysis:

Gene set enrichment analysis was carried out to identify the significantly enriched pathways associated with the potential candidate biomarkers (Weiqing Yang). The genes were first ranked on the basis of

DESeq2 wald statistics, p adjusted value and normalized enrichment scores. KEGG and Gene ontology terms were used for mapping out the GSEA pathways. Gene ontology significantly enriched pathways all had negative enrichment scores meaning that the pathways were all associated with downregulating genes. The most notable ones were vesicle cargo loading, regulation of autophagosome assembly, regulation of vacuole organization and post-Golgi vesicle-mediated transport. Some pathways also had positive enrichment scores like spinal cord patterning, oxygen transport and mucosal immune response. KEGG based GSEA highlighted pathways like autophagy, endocytosis, natural killer cell-mediated cytotoxicity (Alex M Abel), B-cell receptor signalling pathway (Dimitar G Efremov) and maturity onset diabetes of the young (as shown in figure 28).

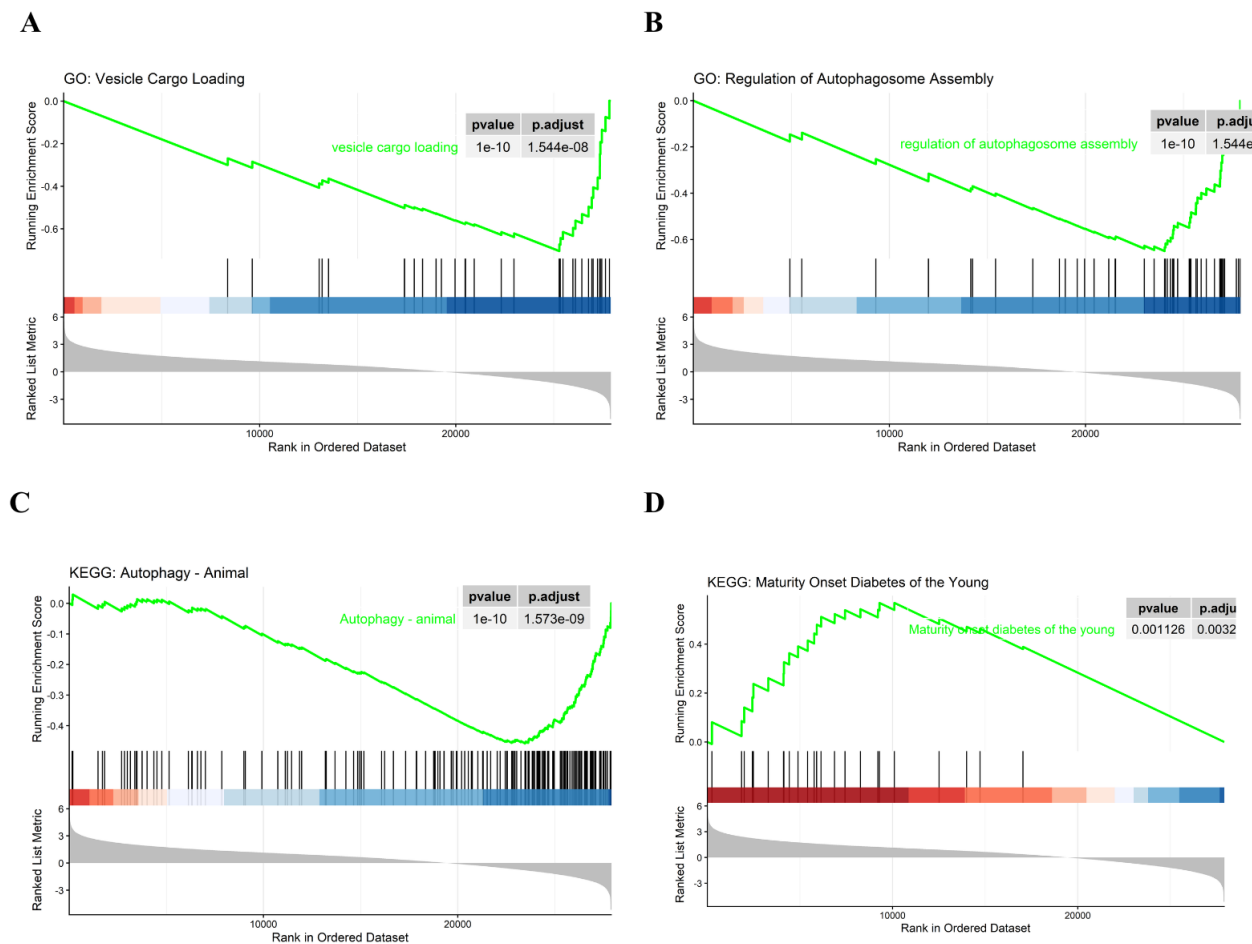


Figure 28: Panelled image depicting the top significantly enriched GO-GSEA pathways (A&B); and top significantly enriched KEGG GSEA pathways (C&D)

4.11 Tissue expression analysis:

Using GTEx database, the expression of the hub genes across human tissues was done (shown in figure 29). AMPD2 had high expression in adrenal gland, spleen, lung, liver and adipose tissue. ABCG1 was highly expressed in blood associated tissue while the gene HBA2 was seen to be highly expressed in whole blood.

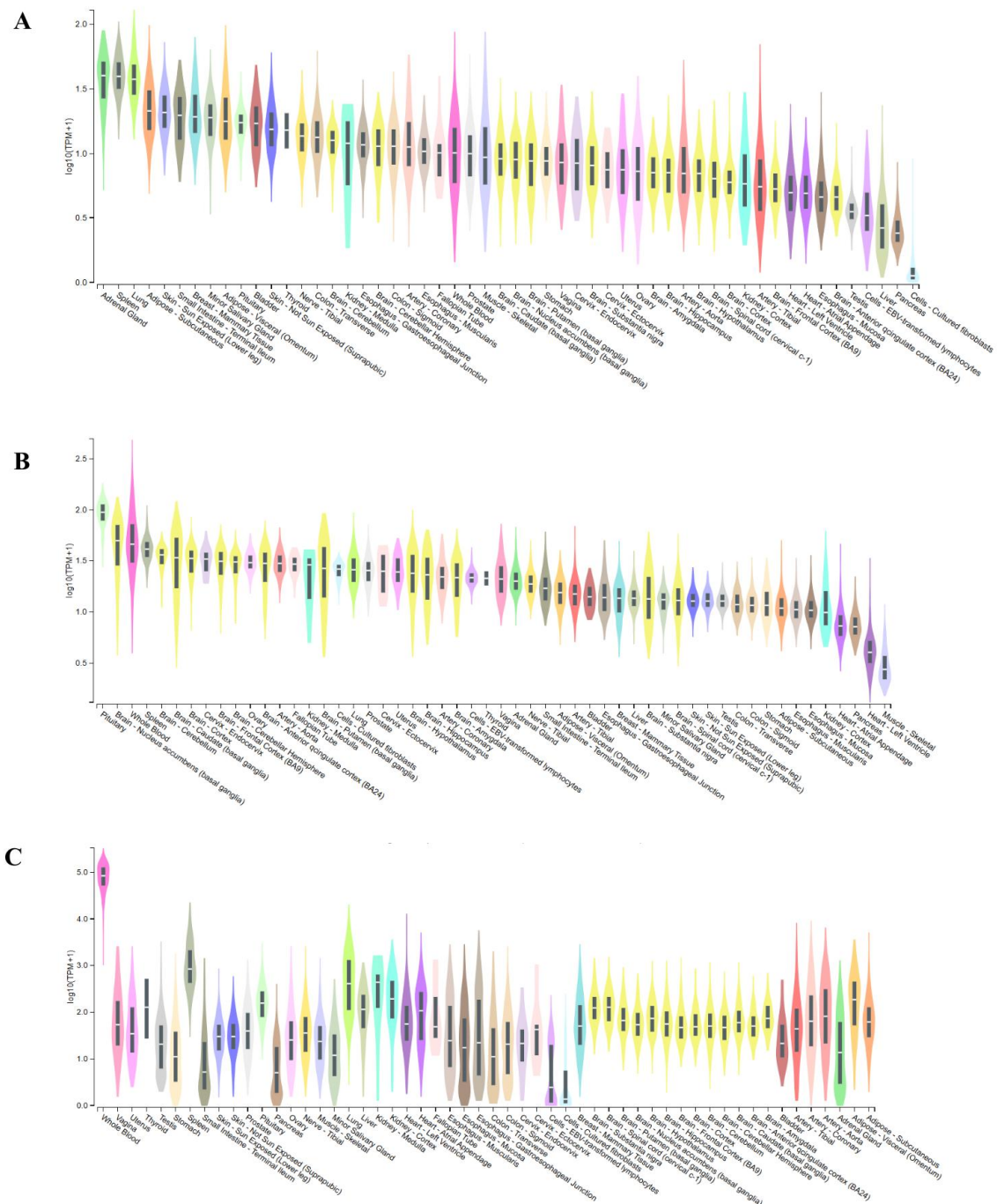


Figure 29: Figure depicting gene expression analysis of (A) ABCG1; (B) AMPD2; and (C) HBA2 across all the tissues

4.12 Molecular Interaction Analysis:

Using AutoDock Vina, interaction occurring between bisphenol A and the candidate biomarkers was done. 3-dimensional structures for ABCG1, AMPD2 and HBA2 were retrieved from PDB (with PDB ID's 7R8D, 4NO3 and 1A01 respectively). Canonical SMILE structure of bisphenol A was retrieved from PubChem. The complex of AMPD2 and BPA yielded a binding affinity of -7.4 kcal/mol (shown in figures 30 and 31). While the complex of BPA and ABCG1 and HBA2 yielded a binding affinity of -8.1 respectively.

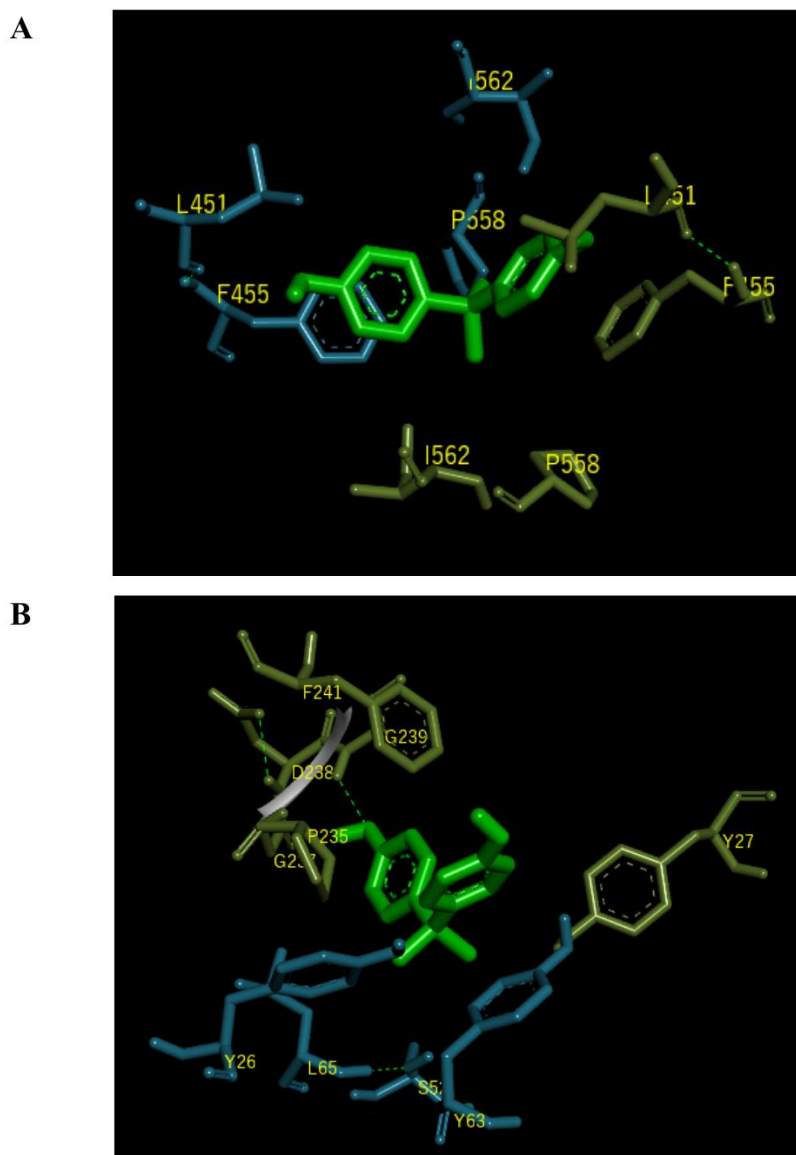


Figure 30: Complex formed between Bisphenol A and (A) Adenosine monophosphate deaminase 2; (B) ATP Binding Cassette sub family G member 1

C

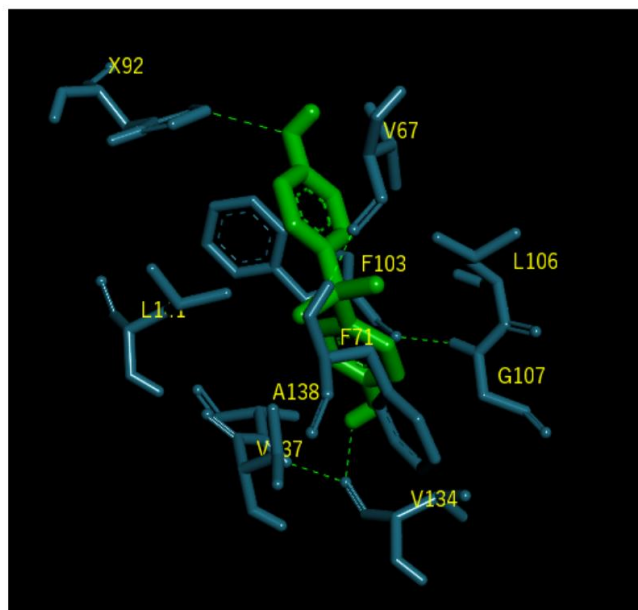


Figure 31: Complex formed between Bisphenol A and (C) Haemoglobin Subunit Alpha 2

CONCLUSION

This study used transcriptomic and machine learning methodologies to evaluate the connection between bisphenol-A and type 2 diabetes (Yan Zhang). Differential expression analysis helped in identifying 355 dysregulated genes which were compared against intersection of BPA-T2D genes to find 52 overlapping elements. Followed by which machine learning based hub gene identification was done using SVM, RF and LASSO; followed by ROC curve analysis which lead to the identification of potential candidate biomarkers ABCG1, AMPD2 and HBA2. Lastly docking studies and functional enrichment was done to identify the role of these candidate genes in individuals suffering from type 2 diabetes. Adenosine Monophosphate Deaminase 2 (AMPD2) is an enzyme involved in energy and glucose metabolism (Athanasius Wrin Hudoyo). Mouse model studies have identified that done on mice with and without AMPD2 highlighted that AMPD2 can effectively regulate gluconeogenesis through the action of specific enzymes and absence of AMPD2 leads to low amounts of glucose being produced (Athanasius Wrin Hudoyo). This study took another parameter in consideration as the mice were given high fructose diet which showed that downregulation of AMPD2 protects against the glucose metabolism due to consumption of high levels of fructose (Athanasius Wrin Hudoyo). AMPD2 is involved in AMPK activity, gluconeogenesis, glucose tolerance, and hepatic lipid metabolism through literature study (Athanasius Wrin Hudoyo). Functional enrichment analysis of AMPD2 also related to pathways like insulin signaling, mTOR signaling, autophagy, sphingolipid signaling, and lipid metabolism which are closely associated with glucose homeostasis and metabolic regulation; factors involved in pathogenesis of T2D (Scherthaner.). HBA2 gene is involved in oxygen transport and erythrocyte function, under the state of hyperglycaemia alterations in the levels of haemoglobin are observed (Joseph Malaba). A scientific study done with case and control samples highlighted that individuals with diabetes had a somewhat different blood composition (Joseph Malaba). This can be used for monitoring or assessing diabetes as under hyperglycaemia this altered composition is observed (Joseph Malaba). Another study highlighted the role of ABCG1 in diabetes (Yinxia Su), it showed that methylation of this gene increased susceptibility of diabetes in individuals (Yinxia Su). It also showed

comorbidity with conditions like obesity, dyslipidaemia and hypertension (Yinxia Su). The gene ABCG1 is involved in cholesterol efflux, lipid homeostasis and insulin secretion (Yinxia Su). This helps us in stating that exposure to BPA can induce or drive epigenetic modification which further promote disease progression (Yinxia Su). All these findings sound promising and form a relation between BPA and Type 2 diabetes, supported with the docking studies which yielded binding interactions of -7.4 and - 8.1 kcal/mol respectively. However, the exact mechanism still remains elusive. And the findings of this study need robust wet lab validation in order to concretely claim that the identified hub genes can act as diagnostic biomarkers and what exact role do they play in BPA induced metabolic disease progression.

REFERENCES

1. Ada Hamosh, Alan F Scott , Joanna S Amberger , Carol A Bocchini and Victor A McKusick. "Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders." *Nucleic Acids Research* (December 2004): D514 - D517.
2. Alberto Santos , Kalliopi Tsafou , Christian Stolte , Sune Pletscher-Frankild , Seán I. O'Donoghue and Lars Juhl Jensen. "Comprehensive comparison of large-scale tissue expression datasets." *PeerJ* (June 2015).
3. Aleksandr A Tkachenko, Ziravard N Tonyan , Yulia A Nasykhova , Yury A Barbitoff , Iaroslav N Renev , Maria M Danilova , Anastasiia A Basipova , Olga B Glavnova , Dmitrii E Polev , Sergey V Chepanov , Sergey A Selkov , Nikita V Golovkin. "Cross-Study Meta-Analysis of Blood Transcriptomes in Type 2 Diabetes." *International Journal of Molecular Sciences* (December 2025).
4. Alex M Abel, Chao Yang , Monica S Thakar and Subramaniam Malarkannan. "Natural Killer Cells: Development, Maturation, and Clinical Utilization." *Frontiers in Immunology*. (August 2018).
5. Alice Marmugi, Simon Ducheix, Frédéric Lasserre, Arnaud Polizzi, Alain Paris, Nathalie Priymenko, Justine Bertrand-Michel, Thierry Pineau, Hervé Guillou, Pascal G.P. Martin, Laïla Mselli-Lakhal. "Low doses of bisphenol a induce gene expression related to lipid synthesis and trigger triglyceride accumulation in adult mouse liver." *Heptaology* (September 2011): 395 - 407.
6. Allan Peter Davis, Cynthia J Grondin, Robin J Johnson, Daniela Sciaky, Jolene Wieggers, Thomas C Wieggers, and Carolyn J Mattingly. "Comparative Toxicogenomics Database (CTD): update 2021." *Nucleic Acids Research*. (2021).
7. Allan Peter Davis, Cynthia J. Grondin , Kelley Lennon-Hopkins , Cynthia Saraceni-Richards , Daniela Sciaky , Benjamin L. King , Thomas C. Wieggers , Carolyn J. Mattingly. "The Comparative Toxicogenomics Database's 10th year anniversary: update 2015." *Nucleic Acids Research* (January 2015): D914 - D920.
8. Allende., Ramón Alonso. "Accelerating searches of research grants and scientific literature with novo|seekSM." *Nature Methods*. (May 2009): i - ii.
9. Anne M Minihane, Sophie Vinoy , Wendy R Russell , Athanasia Baka , Helen M Roche , Kieran M Tuohy , Jessica L Teeling , Ellen E Blaak , Michael Fenech , David Vauzour , Harry J McArdle , Bas H A Kremer , Luc

- Sterkman , Katerina Vafeiadou. "Low-grade inflammation, diet composition and health: current research evidence and its translation." *British Journal of Nutrition*. (July 2015): 999 - 1012.
10. Antoine Daina, Olivier Michielin , Vincent Zoete. "SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules." *Nucleic Acids Research* (May 2019): W357 - W364.
 11. Athanasius Wrin Hudoyo, Tetsuaki Hirase , Andreas Tandelillin , Masahiko Honda , Manabu Shirai , Jidong Cheng , Hiroko Morisaki , Takayuki Morisaki. "Role of AMPD2 in impaired glucose tolerance induced by high fructose diet." *July 2017* (Molecular Genetic and Metabolism Reports.): 23 - 29.
 12. Biovia., Discovery Studio. "Discovery Studio Modelling Environment, Dassault Systems." (2015).
 13. Bo Wu, Xiao-hong Lan, Xu-qing Chen, Qin-yan Wu, Zhi-yang Yu, Yue-kun Wang, Yang Yang, Xiao-chen Cao, Xing-dong Wang, Shuang-yan Zhou and Ling Lin. "Integrating network toxicology, machine learning, and molecular dynamics simulations to reveal tanshinone iia's dual mechanisms in TNBC and doxorubicin-induced cardiotoxicity." *Scientific Reports* (February 2026).
 14. Bobak., Carly A. "Ab2olish Tb: Analysis of Blood and Breath'Omics to Lend Insights and Strategies for Hindering Tuberculosis." Doctoral Dissertation. 2021.
 15. Breiman., Leo. "Random Forests." *Machine Learning* (October 2001): 5 - 32.
 16. Butler, Alexandra E., et al. " β -Cell Deficit and Increased β -Cell Apoptosis in Humans With Type 2 Diabetes." *Diabetes* (January 2003): 102 - 110.
 17. C Weyer, T Funahashi , S Tanaka , K Hotta , Y Matsuzawa , R E Pratley and P A Tataranni. "Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia." *The Journal of Clinical Endocrinology and Metabolism* (May 2001): 1930 - 1935.
 18. Carl F. Schaefer, Kira Anthony , Shiva Krupa , Jeffrey Buchoff , Matthew Day , Timo Hannay , Kenneth H. Buetow. "PID: the Pathway Interaction Database." *Nucleic Acids Research* (January 2009): D674 - D679.
 19. Cerf., Marlon E. "Beta Cell Dysfunction and Insulin Resistance." *Frontiers in Endocrinology* (March 2013).
 20. Chenglin Wang, Xin Wang, Yaxing Deng, Yingchun Hu and Li Hu. "Network toxicology combined with molecular docking technology to explore the molecular mechanism of amatoxin causing liver injury." *Scientific Reports* (July 2025).

21. Chiarelli, Eleonora Rotondo and Francesco. "Endocrine-Disrupting Chemicals and Insulin Resistance in Children." *Biomedicines* (May 2020).
22. Clare Eckold, Vinod Kumar , January Weiner , Bacht Alisjahbana , Anca-Lelia Riza , Katharina Ronacher , Jorge Coronel , Sarah Kerry-Barnard , Stephanus T Malherbe , Leanie Kleynhans , Kim Stanley , Rovina Ruslami , Mihai Ioana , Cesar Ugarte-Gil. "Impact of Intermediate Hyperglycemia and Diabetes on Immune Dysfunction in Tuberculosis." *Clinical Infectious Diseases* (June 2020): 69 - 78.
23. Conrad L Schoch, Stacy Ciufu , Mikhail Domrachev , Carol L Hotton , Sivakumar Kannan , Rogneda Khovanskaya , Detlef Leipe , Richard Mcveigh , Kathleen O'Neill , Barbara Robbertse , Shobha Sharma , Vladimir Soussov , John P Sullivan , Lu Sun, Seán Turner. "NCBI Taxonomy: a comprehensive update on curation, resources and tools. ." *Database*. (August 2020).
24. Consortium., The Gene Ontology. "Gene Ontology Consortium: going forward." *Nucleic Acids Research*. (January 2015): D1049–D1056.
25. Consortium., The UniProt. "UniProt: the universal protein knowledgebase." *Nucleic Acids Research* (February 2018): D158 - D169.
26. Damian Szklarczyk, Alberto Santos , Christian von Mering , Lars Juhl Jensen , Peer Bork , Michael Kuhn. "STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data." *Nucleic Acids Research*. (January 2016): D380 - D384.
27. David Croft, Antonio Fabregat Mundo , Robin Haw , Marija Milacic , Joel Weiser , Guanming Wu , Michael Caudy , Phani Garapati , Marc Gillespie , Maulik R. Kamdar , Bijay Jassal , Steven Jupe , Lisa Matthews , Bruce May and Stanislav Palatnik. "The Reactome pathway knowledgebase." *Nucleic Acids Research* (January 2014): D472 - D477.
28. David Moher, Alessandro Liberati , Jennifer Tetzlaff , Douglas G Altman , The PRISMA Group. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLoS Medicine* (July 2009).
29. David Ryan Koes, Matthew P. Baumgartner and Carlos J. Camacho. "Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise." *Journal of Chemical Information and Modeling* (February 2013): 1893 - 1904.

30. Dimitar G Efremov, Sven Turkalj and Luca Laurenti. "Mechanisms of B Cell Receptor Activation and Responses to B Cell Receptor Inhibitors in B Cell Malignancies." *Cancers*. (May 2020).
31. Dong., Peter Thomas and Jing. "Binding and activation of the seven-transmembrane estrogen receptor GPR30 by environmental estrogens: A potential novel mechanism of endocrine disruption." *The Journal of Steroid Biochemistry and Molecular Biology*. (December 2006): 175 - 179.
32. Eva Tudurí, Laura Marroqui , Reinaldo S Dos Santos , Iván Quesada , Esther Fuentes , Paloma Alonso-Magdalena. "Timing of Exposure and Bisphenol-A: Implications for Diabetes Development." *Frontiers in Endocrinology*. (October 2018).
33. Fang Ren, Jing Wei , Qingxin Chen , Mengling Hu , Lu Yu , Jianing Mi , Xiaogang Zhou , Dalian Qin , Jianming Wu , Anguo Wu. "Artificial intelligence-driven multi-omics approaches in Alzheimer's disease: Progress, challenges, and future directions." *Acta Pharmaceutica Sinica. B* (July 2025): 4327 - 4385.
34. Fanny Rancière, Jasmine G. Lyons, Venurs H.Y. Loh, Jérémie Botton, Tamara Galloway, Tiange Wang, Jonathan E. Shaw and Dianna J. Magliano. "Bisphenol A and the risk of cardiometabolic disorders: a systematic review with meta-analysis of the epidemiological evidence." *Environmental Health* (May 2015).
35. Federica Cariati, Luigi Carbone , Alessandro Conforti , Francesca Bagnulo , Stefania Ramona Peluso , Consolata Carotenuto , Cira Buonfantino , Erminia Alviggi , Carlo Alviggi , Ida Strina. "Bisphenol A-Induced Epigenetic Changes and Its Effects on the Male Reproductive System." *Frontiers in Endocrinology* (July 2020).
36. Feng Zhu, Zhe Shi , Chu Qin , Lin Tao , Xin Liu , Feng Xu , Li Zhang , Yang Song , Xianghui Liu , Jingxian Zhang , Bucong Han , Peng Zhang , Yuzong Chen. "Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery." *Nucleic Acids Research* (January 2012): D1128 - D1136.
37. Francesca Farrugia, Alexia Aquilina , Josanne Vassallo , Nikolai Paul Pace. "Bisphenol A and Type 2 Diabetes Mellitus: A Review of Epidemiologic, Functional, and Early Life Factors." *International Journal of Environmental Research and Public Health*. (January 2021).
38. Frida Belinky, Noam Nativ, Gil Stelzer, Shahar Zimmerman, Tsippi Iny Stein, Marilyn Safran and Doron Lancet. "PathCards: multi-source consolidation of human biological pathways." *Database*. (February 2015).

39. Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell and Arthur J. Olson. "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." *Journal of Computational Chemistry* (April 2009): 2785 - 2791.
40. Gil Stelzer, Aron Inger, Tsviya Olender, Tsippi Iny-Stein, Irina Dalah, Arye Harel, Marilyn Safran, and Doron Lancet. "GeneDecks: Paralog Hunting and Gene-Set Distillation with GeneCards Annotation." *OMICS: A Journal of Integrative Biology*. (December 2009): 477 - 487.
41. Gil Stelzer, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, Ron Nudel, Iris Lieder, Yaron Mazor, Sergey Kaplan, Dvir Dahary, David Warshawsky, Yaron Golan, Asher Kohn, Noa Rappaport, Marilyn Safran. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." *Current Protocols in Bioinformatics*. (June 2016): 1.30.1 - 1.30.33.
42. Hala F S Abulehia, Noor Shafina Mohd Nor, Siti Hamimah Sheikh Abdul Kadir. "The Current Findings on the Impact of Prenatal BPA Exposure on Metabolic Parameters: In Vivo and Epidemiological Evidence." *Nutrients* (July 2022).
43. Hamiltom, James P. "Epigenetics: Principles and Practice." *Digestive Diseases* (July 2011): 130 - 135.
44. Harner., Tijana Vasiljevic and Tom. "Bisphenol A and its analogues in outdoor and indoor air: Properties, sources and global levels." *Science of the Total Environment*. (October 2021).
45. Heewon Song, Joonwoo Park , Phuong T.C Bui , KeunOh Choi , Myung Chan Gye , Yun-Chul Hong , Jin Hee Kim and Young Joo Lee. "Bisphenol A induces COX-2 through the mitogen-activated protein kinase pathway and is associated with levels of inflammation-related markers in elderly populations." *Environmental Research* (October 2017): 490 - 498.
46. Helen Berman, Kim Henrick and Haruki Nakamura. "Announcing the worldwide Protein Data Bank." *Nature Structural & Molecular Biology*. (December 2003).
47. Helen M. Berman, John Westbrook , Zukang Feng , Gary Gilliland , T. N. Bhat , Helge Weissig , Ilya N. Shindyalov , Philip E. Bourne. "The Protein Data Bank. ." *Nucleic Acids Research* (January 2000): 235 - 242.
48. Hung-Hsin Chen, Lauren E Petty , Kari E North , Joseph B McCormick , Susan P Fisher-Hoch , Eric R Gamazon , Jennifer E Below. "Novel diabetes gene discovery through comprehensive characterization and integrative analysis of longitudinal gene expression changes." *Human Molecular Genetics* (February 2022): 3191 - 3205.

49. Hye-Rim Lee, Tae-Hee Kim and Kyung-Chul Choi. "Functions and physiological roles of two types of estrogen receptors, ER α and ER β , identified by estrogen receptor knockout mouse." *Laboratory Animal Research* (June 2012): 71 - 76.
50. Iain A. Lang, Tamara S. Galloway and Alan Scarlett. "Association of Urinary Bisphenol A Concentration With Medical Disorders and Laboratory Abnormalities in Adults." *JAMA Network* (September 2008): 1303 - 1310.
51. Ilaria Cimmino, Francesca Fiory , Giuseppe Perruolo , Claudia Miele , Francesco Beguinot , Pietro Formisano , Francesco Oriente. "Potential Mechanisms of Bisphenol A (BPA) Contributing to Human Disease." *International Journal of Molecular Sciences* (August 2020).
52. Iqra Hameed, Shariq R Masoodi , Shahnaz A Mir , Mudasar Nabi , Khalid Ghazanfar and Bashir A Ganai. "Type 2 diabetes mellitus: From a metabolic disorder to an inflammatory condition." *World Journal of Diabetes* (May 2015): 598 - 612.
53. Jangwoo Lee, Kyungho Choi, Jeongim Park, Hyo-Bang Moon, Gyuyeon Choi, Jeong Jae Lee, Eunsook Suh, Hai-joong Kim, So-Hee Eun, Gun-Ha Kim, Geum Joon Cho, Sung Koo Kim, Sungjoo Kim, Su Young Kim, Seunghyo Kim, Soyong Eom, Sooran Choi, Young Don Kim. "Bisphenol A distribution in serum, urine, placenta, breast milk, and umbilical cord serum in a birth panel of mother–neonate pairs." *Science of the Total Environment* (June 2018): 1494 - 1501.
54. Jie Wei, Yi Lin , Yuanyuan Li , Chenjiang Ying , Jun Chen , Liqiong Song , Zhao Zhou , Ziquan Lv , Wei Xia , Xi Chen , Shunqing Xu. "Perinatal Exposure to Bisphenol A at Reference Dose Predisposes Offspring to Metabolic Syndrome in Adult Rats on a High-Fat Diet." *Endocrinology* (August 2011): 3049 - 3061.
55. Jin Hee Kim, Yoo Hee Cho and Yun-Chul Hong. "MicroRNA expression in response to bisphenol A is associated with high blood pressure." *Environmental International* (May 2020).
56. Jonathan G Boucher, Rémi Gagné , Andrea Rowan-Carroll , Adèle Boudreau , Carole L Yauk , Ella Atlas. "Bisphenol A and Bisphenol S Induce Distinct Transcriptional Profiles in Differentiating Human Primary Preadipocytes." *PloS One* (September 2016).
57. Jorge Enrique González-Casanova, Valmore Bermúdez , Nelson Javier Caro Fuentes , Lissé Chiquinquirá Angarita , Nelson Hernando Caicedo , Jocelyn Rivas Muñoz and Diana Marcela Rojas-Gómez. "New Evidence

on BPA's Role in Adipose Tissue Development of Proinflammatory Processes and Its Relationship with Obesity." *International Journal of Molecular Sciences* (May 2023).

58. Joseph Malaba, Paul Kosiyo and Bernard Guyah. "Haemoglobin types and variant interference with HbA1c and its association with uncontrolled HbA1c in type 2 diabetes mellitus." *BMC Research Notes*. (November 2024).
59. Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstern, David A. Evans, Chia-Chun Hung, Michael O'Neill and David Reiman. "Accurate structure prediction of biomolecular interactions with AlphaFold 3." *Nature* (May 2024): 493 - 500.
60. Kelly K Ferguson, David E Cantonwine , Thomas F McElrath , Bhramar Mukherjee and John D Meeker. "Repeated measures analysis of associations between urinary bisphenol-A concentrations and biomarkers of inflammation and oxidative stress in pregnancy." *Reproductive Toxicology* (October 2016): 93 - 98.
61. La-or Chailurkit, Pechngam Tengpraettanakorn and Boonsong Ongphiphadhanakul. "Is bisphenol A exposure associated with the development of glucose intolerance and increased insulin resistance in Thais?" *Nutrition and Health* (May 2018).
62. Lieder., Shani Ben-Ari Fuchs and Iris. "GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data." *OMICS: A Journal of Integrative Biology*. (March 2016): 139 - 151.
63. Linlin Song, Hongli Zhang and Wang Yang. "Multiple machine learning algorithms identified SLC6A8 as a diagnostic biomarker of the late stage of Hepatocellular carcinoma." *Discover Oncology* (April 2025).
64. M Rebhan, V Chalifa-Caspi , J Prilusky and D Lancet. "GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support." *Bioinformatics*. (January 1998): 656 - 664.
65. M. Stuart Armstrong, Garrett M. Morris, Paul W. Finn, Raman Sharma, Loris Moretti, Richard I. Cooper and W. Graham Richards. "ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics." *Journal of Computer-Aided Molecular Design* (July 2010): 789 - 801.
66. M. Stuart Armstrong, Paul W. Finn, Garrett M. Morris and W. Graham Richards. "Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension." *Journal of Computer-Aided Molecular Design* (August 2011): 785 - 790.

67. Maggie S Burhans, Derek K Hagman , Jessica N Kuzma , Kelsey A Schmidt , Mario Kratz. "Contribution of adipose tissue inflammation to the development of type 2 diabetes mellitus." *Comprehensive Physiology* (December 2018): 1 - 58.
68. Maria K. Björnsdotter, Jacob de Boer and Ana Ballesteros-Gómez. "Bisphenol A and replacements in thermal paper: A review." *Chemosphere* (September 2017): 691 - 706.
69. Marilyn Safran, Irina Dalah , Justin Alexander , Naomi Rosen , Tsippi Iny Stein , Michael Shmoish , Noam Nativ , Iris Bahir , Tirza Doniger , Hagit Krug , Alexandra Sirota-Madi , Tsviya Olender , Yaron Golan and Gil Stelzer. "GeneCards Version 3: the human gene integrator." *Database*. (August 2010).
70. Marilyn Safran, Irina Solomon , Orit Shmueli , Michal Lapidot , Shai Shen-Orr , Avital Adato , Uri Ben-Dor , Nir Esterman , Naomi Rosen , Inga Peter , Tsviya Olender , Vered Chalifa-Caspi and Doron Lancet. "GeneCards™ 2002: towards a complete, object-oriented, human gene compendium." *Bioinformatics*. (November 2002): 1542 - 1543.
71. Marilyn Safran, Vered Chalifa-Caspi , Orit Shmueli , Tsviya Olender , Michal Lapidot , Naomi Rosen , Michael Shmoish , Yakov Peter , Gustavo Glusman , Ester Feldmesser , Avital Adato , Inga Peter , Miriam Khen , Tal Atarot , Yoram Groner , Doron Lancet. "Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE." *Nucleic Acids Research* (January 2003): 142 - 146.
72. Maxim Shklar, Liora Strichman-Almashanu , Orit Shmueli , Michael Shmoish , Marilyn Safran , Doron Lancet. "GeneTide—Terra Incognita Discovery Endeavor: a new transcriptome focused member of the GeneCards/GeneNote suite of databases." *Nucleic Acids Research*. (January 2005): D556 - D561.
73. Mei-Ling Huang, Yung-Hsiang Hung, W M Lee, R K Li and Bo-Ru Jiang. "SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier." *The Scientific World Journal*. (September 2014).
74. Melissa J. Landrum, Jennifer M. Lee, Mark Benson, Garth Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, Jennifer Hart, Douglas Hoffman, Jeffrey Hoover, Wonhee Jang, Kenneth Katz, Michael Ovetsky, George Riley and Amanjeev Sethi. "ClinVar: public archive of interpretations of clinically relevant variants." *Nucleic Acids Research* (November 2015): D862 - D868.

75. Michael J Keiser, Bryan L Roth, Blaine N Armbruster, Paul Ernsberger, John J Irwin and Brian K Shoichet. "Relating protein pharmacology by ligand chemistry." *Nature Biotechnology*. (February 2007): 197 - 206.
76. Mihaly Varadi, Damian Bertoni , Paulyna Magana , Urmila Paramval , Ivanna Pidruchna , Malarvizhi Radhakrishnan , Maxim Tsenkov , Sreenath Nair , Milot Mirdita , Jingi Yeo , Oleg Kovalevskiy , Kathryn Tunyasuvunakool , Agata Laydon , Augustin Židek. "AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences." *Nucleic Acids Research* (November 2023): D368 - D375.
77. Minoru Kanehisa, Susumu Goto , Yoko Sato , Masayuki Kawashima , Miho Furumichi , Mao Tanabe. "Data, information, knowledge and principle: back to metabolism in KEGG. ." *Nucleic Acids Research*. (January 2014): D199 - D205.
78. Mónica García García, Yolanda Picó and María Morales-Suárez-Varela. "Effects of Bisphenol A on the Risk of Developing Obesity." *Nutrients* (October 2024).
79. Monica K. Silver, Marie S. O'Neill, Mary Fran R. Sowers and Sung Kyun Park. "Urinary Bisphenol A and Type-2 Diabetes in U.S. Adults: Data from NHANES 2003-2008." *PLOS One* (October 2011).
80. Monica M. Vasquez, Chengcheng Hu, Denise J. Roe, Zhao Chen, Marilyn Halonen and Stefano Guerra. "Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application." *BMC Medical Research Methodology*. (November 2016).
81. Moore., Latarsha J. Carithers and Helen M. "The Genotype-Tissue Expression (GTEx) Project." *Biopreservation and Biobanking*. (2015).
82. Morihiro Matsuda, Iichiro Shimomura, Masataka Sata, Yukio Arita, Makoto Nishida, Norikazu Maeda, Masahiro Kumada, Yoshihisa Okamoto, Hiroyuki Nagaretani, Hitoshi Nishizawa, Ken Kishida, Ryutaro Komuro, Noriyuki Ouchi, Shinji Kihara, Ryozi Nagai. "Role of Adiponectin in Preventing Vascular Stenosis: THE MISSING LINK OF ADIPO-VASCULAR AXIS." *Journal of Biological Chemistry* (October 2002): 37487 - 37491.
83. Nafisa M. Hassan, Amr A. Alhossary, Yuguang Mu and Chee-Keong Kwoh. "Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration." *Scientific Reports* (November 2017).

84. Naomi Rosen, Vered Chalifa-Caspi, Orit Shmueli, Avital Adato, Michal Lapidot, Julie Stampnitzky, Marilyn Safran and Doron Lancet. "GeneLoc: exon-based integration of human genome maps." *Bioinformatics* (February 2003): i222 - i224.
85. Noa Rappaport, Michal Twik, Inbar Plaschkes, Ron Nudel, Tsippi Iny Stein, Jacob Levitt, Moran Gershoni, C. Paul Morrey, Marilyn Safran, Doron Lancet. "MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search." *Nucleic Acids Research* (January 2017): D877 - D887.
86. Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch and Geoffrey R Hutchison. "Open Babel: An open chemical toolbox." *Journal of Cheminformatics*. (October 2011).
87. Olson., Oleg Trott and Arthur J. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading." *Journal of Computational Chemistry* (January 2010): 455 - 461.
88. Olson., Oleg Trott and Arthur J. "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." *Journal of Computational Chemistry* (June 2009): 455 - 461.
89. Orit Shmueli, Shirley Horn-Saban , Vered Chalifa-Caspi , Michael Shmoish , Ron Ophir , Hila Benjamin-Rodrig , Marilyn Safran , Eystan Domany , Doron Lancet. "GeneNote: whole genome expression profiles in normal human tissues." *Comptes Rendus Biologies*. (November 2003): 1067 - 1072.
90. Paloma Alonso-Magdalena, Ana B Ropero , M Pilar Carrera , Christopher R Cederroth , Mathurin Baquié , Benoit R Gauthier , Serge Nef , Enrico Stefani , Angel Nadal. "Pancreatic Insulin Content Regulation by the Estrogen Receptor ER α ." *PLoS One* (April 2008).
91. Paloma Alonso-Magdalena, Elaine Vieira , Sergi Soriano , Lorena Menes , Deborah Burks , Ivan Quesada and Angel Nadal. "Bisphenol A Exposure during Pregnancy Disrupts Glucose Homeostasis in Mothers and Adult Male Offspring." *Environmental Health Perspectives* (May 2010): 1243 - 1250.
92. Paloma Alonso-Magdalena, Francisco J Rivera and Carlos Guerrero-Bosagna. "Bisphenol-A and metabolic diseases: epigenetic, developmental and transgenerational basis." *Environmental Epigenetics* (October 2016).

93. Paloma Alonso-Magdalena, Ivan Quesada and Angel Nadal. "Endocrine disruptors in the etiology of type 2 diabetes mellitus." *Nature Reviews Endocrinology* (April 2011): 346 - 353.
94. Paloma Alonso-Magdalena, Marta García-Arévalo , Iván Quesada and Ángel Nadal. "Bisphenol-A Treatment During Pregnancy in Mice: A New Window of Susceptibility for the Development of Diabetes in Mothers Later in Life." *Endocrinology*. (May 2015): 1656 - 1670.
95. Paloma Alonso-Magdalena, Sergi Sorino, Marta Gracia, Cristina Ripoll, Esther Funetes, Angel Nadal. "Bisphenol-A acts as a potent estrogen via non-classical estrogen triggered pathways." *Molecular and Cellular Endocrinology* (December 2011).
96. Paloma Alonso-Magdalena, Sumiko Morimoto , Cristina Ripoll , Esther Fuentes , Angel Nadal. "The Estrogenic Effect of Bisphenol A Disrupts Pancreatic β -Cell Function In Vivo and Induces Insulin Resistance." *Environmental Health Perspectives* (September 2005): 106 - 112.
97. Patricia M. Bock, Gabriela H. Telo, Rafaela Ramalho, Mariana Sbaraini, Gabriel Leivas, Andreza F. Martins and Beatriz D. Schaan. "The effect of probiotics, prebiotics or synbiotics on metabolic outcomes in individuals with diabetes: a systematic review and meta-analysis." *Diabetologia* (October 2020): 26 - 41.
98. Pelin Yaşar, Gamze Ayaz, Sırma Damla User, Gizem Güpür, Mesut Muyan. "Molecular mechanism of estrogen–estrogen receptor signaling." *Reproductive Medicine and Biology* (December 2016): 4 - 20.
99. Prince., Doveit Antony Charles and Sabrina Evan. "Deciphering the molecular mechanism of NLRP3 in BPA-mediated toxicity: Implications for targeted therapies." *Heliyon* (April 2024).
100. Racine., Jeffrey S. "RStudio: A Platform-Independent IDE for R and Sweave." *Journal of Applied Econometrics*. (October 2011): 167 - 172.
101. Rasko Leinonen, Federico Garcia Diez , David Binns , Wolfgang Fleischmann , Rodrigo Lopez , Rolf Apweiler. "UniProt archive." *Bioinformatics* (March 2004): 3236 - 3237.
102. Rasko Leinonen, Hideaki Sugawara , Martin Shumway on behalf of the International Nucleotide Sequence Database Collaboration. "The Sequence Read Archive." *Nucleic Acids Research* (January 2011): 19 - 21.

103. Robert Petryszak, Tony Burdett , Benedetto Fiorelli , Nuno A. Fonseca , Mar Gonzalez-Porta , Emma Hastings , Wolfgang Huber , Simon Jupp , Maria Keays , Nataliya Kryvych , Julie McMurry , John C. Marioni , James Malone and Karine Megy. "Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments." *Nucleic Acids Research* (January 2014): D926 - D932.
104. Ron Caspi, Tomer Altman , Richard Billington , Kate Dreher , Hartmut Foerster , Carol A. Fulcher , Timothy A. Holland , Ingrid M. Keseler , Anamika Kothari , Aya Kubo , Markus Krummenacker , Mario Latendresse , Lukas A. Mueller and Quang Ong. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic Acids Research* (January 2014): D459 - D471.
105. Ron Edgar, Michael Domrachev and Alex E Lash. "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." *Nucleic Acids Research* (January 2002): 207 - 210.
106. Rorsmam., Frances M. Ashcroft and Patrik. "Diabetes Mellitus and the β Cell: The Last Ten Years." *Cell* (March 2012): 1160 - 1171.
107. Rosaria Meli, Anna Monnolo , Chiara Annunziata , Claudio Pirozzi , Maria Carmela Ferrante. "Oxidative Stress and BPA Toxicity: An Antioxidant Approach for Male and Female Reproductive Dysfunction." *Antioxidants* (May 2020).
108. Rubin., Beverly S. "Bisphenol A: An endocrine disruptor with widespread exposure and multiple effects." *The Journal of Steroid Biochemistry and Molecular Biology* (October 2011): 27 - 34.
109. S. Kaser, T. Tatarczyk, A. Stadlmayr, C. Ciardi, C. Röss, A. Tschoner, A. Sandhofer, B. Paulweber, C. F. Ebenbichler and J. R. Patsch. "Effect of obesity and insulin sensitivity on adiponectin isoform distribution." *European Journal of Clinical Investigation* (October 2008): 827 - 834.
110. Samuel Murail, Sjoerd J de Vries , Julien Rey , Gautier Moroy and Pierre Tufféry. "SeamDock: An Interactive and Collaborative Online Docking Resource to Assist Small Compound Molecular Docking." *Frontiers in Molecular Biosciences*. (September 2021).
111. Sargis., Margaret C Schulz and Robert M. "Inappropriately sweet: Environmental endocrine-disrupting chemicals and the diabetes pandemic." *Advances in Pharmacology* (June 2021): 419 - 456.

112. Schernthaner., Gerit-Holger Schernthaner and Guntram. "Insulin resistance and inflammation in the early phase of type 2 diabetes: potential for therapeutic intervention." *Scandinavian journal of clinical & laboratory investigation. Supplement.* (2005): 30 - 40.
113. Semi Hwang, Jung-eun Lim , Yoonjeong Choi , Sun Ha Jee. "Bisphenol A exposure and type 2 diabetes mellitus risk: a meta-analysis." *BMC Endocrine Disorders* (November 2018).
114. Sergi Soriano, Paloma Alonso-Magdalena , Marta García-Arévalo , Anna Novials , Sarheed J Muhammed , Albert Salehi , Jan-Ake Gustafsson , Ivan Quesada , Angel Nadal. "Rapid Insulinotropic Action of Low Doses of Bisphenol-A on Mouse and Human Islets of Langerhans: Role of Estrogen Receptor β ." *PLoS One* (Febrary 2012).
115. Shadab Ahmad, Leonardo Jose da Costa Gonzales , Emily H Bowler-Barnett , Daniel L Rice , Minjoon Kim , Supun Wijerathne , Aurélien Luciani , Swaathi Kandasamy , Jie Luo , Xavier Watkins , Edd Turner , Maria J Martin , the UniProt Consortium. "The UniProt website API: facilitating programmatic access to protein knowledge." *Nucleic Acids Research.* (July 2025): W547 - W553.
116. Shaoheng Liang, Qingnan Liang, Rui Chen and Ken Chen. "Stratified Test Accurately Identifies Differentially Expressed Genes Under Batch Effects in Single-Cell Data." *IEEE/ACM Transaction of Computational Biology and Bioinformatics* (December 2021): 2072 - 2079.
117. Shehreen Amjad, Md Saidur Rahman , Myung-Geol Pang. "Role of Antioxidants in Alleviating Bisphenol A Toxicity." *Biomolecules* (July 2020).
118. Simon A. Forbes, David Beare , Prasad Gunasekaran , Kenric Leung , Nidhi Bindal , Harry Boutselakis , Minjie Ding , Sally Bamford , Charlotte Cole , Sari Ward , Chai Yin Kok , Mingming Jia , Tisham De , Jon W. Teague and Michael R. Stratton. "COSMIC: exploring the world's knowledge of somatic mutations in human cancer." *Nucleic Acids Research.* (January 2015): D805 - D811.
119. Soheila Rahmani, Nazila Pour Khalili, Fazullah Khan, Shokoufeh Hassani, Elmira Ghafour-Boroujedi and Mohammad Abdollahi. "Bisphenol A: What lies beneath its induced diabetes and the epigenetic modulation?" *Life Sciences* (October 2018).
120. Stefan Günther, Michael Kuhn , Mathias Dunkel , Monica Campillos , Christian Senger , Evangelia Petsalaki , Jessica Ahmed , Eduardo Garcia Urdiales , Andreas Gewiess , Lars Juhl Jensen , Reinhard Schneider

- , Roman Skoblo and Robert B. Russell. "SuperTarget and Matador: resources for exploring drug-target relationships." *Nucleic Acids Research* (January 2008): D919 - D922.
121. Sune Pletscher-Frankild, Albert Palleja, Kalliopi Tsafou, Janos X. Binder and Lars Juhl Jensen. "DISEASES: Text mining and data integration of disease–gene associations." *Methods*. (March 2015): 83 - 89.
 122. Sunghwan Kim, Jie Chen , Tiejun Cheng , Asta Gindulyte , Jia He , Siqian He , Qingliang Li , Benjamin A Shoemaker , Paul A Thiessen , Bo Yu , Leonid Zaslavsky , Jian Zhang , Evan E Bolton. "PubChem in 2021: new data content and improved web interfaces. ." *Nucleic Acids Research*. (January 2021): D1388 - D1395.
 123. T. Yamauchi, J. Kamon, H. Waki, Y. Terauchi, N. Kubota, K. Hara, Y. Mori, T. Ide, K. Murakami, N. Tsuboyama-Kasaoka, O. Ezaki, Y. Akanuma, O. Gavrilova, C. Vinson, M.L. Reitman, H. Kagechika, K. Shudo, M. Yoda, Y. Nakano, K. Tobe, R. Nagai and S. Kimura. "The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity." *Nature Medicine* (August 2001): 941 - 946.
 124. Tanya Barrett, Stephen E. Wilhite , Pierre Ledoux , Carlos Evangelista , Irene F. Kim , Maxim Tomashevsky , Kimberly A. Marshall , Katherine H. Phillippy , Patti M. Sherman , Michelle Holko , Andrey Yefanov , Hyeseung Lee , Naigong Zhang. "NCBI GEO: archive for functional genomics data sets—update." *Nucleic Acids Research* (January 2013): 991 - 995.
 125. Tetsuya Adachi, Koichiro Yasuda, Chisato Mori, Mariko Yoshinaga, Norihiko Aoki, Gozoh Tsujimoto and Kinsuke Tsuda. "Promoting insulin secretion in pancreatic islets by means of bisphenol A and nonylphenol via intracellular estrogen receptors." *Food and Chemical Toxicology* (May 2005).
 126. Unai Galicia-Garcia, Asier Benito-Vicente , Shifa Jebari , Asier Larrea-Sebal , Haziq Siddiqi , Kepa B Uribe , Helena Ostolaza , César Martín. "Pathophysiology of Type 2 Diabetes Mellitus. ." *International Journal of Molecular Sciences* (August 2020).
 127. Vito., Maurício Krause and Giuseppe De. "Type 1 and Type 2 Diabetes Mellitus: Commonalities, Differences and the Importance of Exercise and Nutrition." *Nutrients* (October 2023).
 128. Vivian Law, Craig Knox , Yannick Djoumbou , Tim Jewison , An Chi Guo , Yifeng Liu , Adam Maciejewski , David Arndt , Michael Wilson , Vanessa Neveu , Alexandra Tang , Geraldine Gabriel , Carol Ly , Sakina Adamjee , Zerihun T. Dame and Beomsoo Han. "DrugBank 4.0: shedding new light on drug metabolism." *Nucleic Acids Research* (January 2014): D1091 - D1097.

129. Waqas Sami, Tahir Ansari , Nadeem Shafique Butt , Mohd Rashid Ab Hamid. "Effect of diet on type 2 diabetes mellitus: A review." *International Journal of Health Sciences* (April 2017): 65 - 71.
130. Wayez Naqvi, Ananya Singh, Prekshi Garg and Prachi Srivastava. "Impact of Prolonged Use of COVID-19 Drugs on the Human Neurological System using Insilco Drug-gene Interaction." *Coronaviruses* (August 2023).
131. Wayez Naqvi, Prekshi Garg and Prachi Srivastava. "Transcriptome Derived Artificial neural networks predict PRRC2A as a potent biomarker for epilepsy." *Journal of Genetic Engineering and Biotechnology*. (May 2025).
132. Weiqing Yang, Sunan Yang, Wanjia Zhang, Qian Zhou, Xiaoqin Tin and Yin Xu. "Exploring the toxicological impact of di(2-ethylhexyl) phthalate exposure on Alzheimer's disease via network toxicology and integrative bioinformatics analysis." *Brain Research Bulletin*. (November 2025).
133. Winn., Megan E. Cull and Louise M. "Bisphenol A and its potential mechanism of action for reproductive toxicity." *Toxicology* (February 2025).
134. Wolfgang Völkel, Thomas Colnot, György A. Csanády, Johannes G. Filser and Wolfgang Dekant. "Metabolism and Kinetics of Bisphenol A in Humans at Low Doses Following Oral Administration." *Chemical Research in Toxicology* (September 2002): 1281 - 1287.
135. Xiaoqiang Wang, Desiree Ha , Ryohei Yoshitake , Yin S Chan , David Sadava , Shiuan Chen. "Exploring the Biological Activity and Mechanism of Xenoestrogens and Phytoestrogens in Cancers: Emerging Methods and Concepts." *International Journal of Molecular Sciences* (August 2021).
136. Xu Hong, Yi Zhou, Zhiyuan Zhu, Yuting Li, Zuo Li, Yuheng Zhang, Xinxin Hu, Fuhai Zhu, Yong Wang, Mingliang Fang, Yichao Huang and Tong Shen. "Environmental endocrine disruptor Bisphenol A induces metabolic derailment and obesity via upregulating IL-17A in adipocytes." *Environmental International* (January 2023).
137. Y Lin, X Sun , L Qiu , J Wei , Q Huang , C Fang 1, T Ye , M Kang , H Shen , S Dong. "Exposure to bisphenol A induces dysfunction of insulin secretion and apoptosis through the damage of mitochondria in rat insulinoma (INS-1) cells." *Cell Death & Disease* (January 2013).

138. Yan Zhang, Shan Lu , Yijun Dong and Jin Shang. "Unveiling the impact of bisphenol a exposure on gene expression and immune response in diabetic nephropathy through integrative toxicogenomics and molecular dynamics approaches." *Diabetology and Metabolic Syndrome* (August 2025).
139. Yasushi Okuno, Akiko Tamon , Hiroaki Yabuuchi , Satoshi Nijima , Yohsuke Minowa , Koichiro Tonomura , Ryo Kunimoto and Chunlai Feng. "GLIDA: GPCR—ligand database for chemical genomics drug discovery—database and tools update. ." *Nucleic Acids Research* (January 2008): D907 - D912.
140. Yinxia Su, Bo Shang, Xiaoyuan Hu, Zhihao Zhang, Li Wang, Kun Luo, Hua Yao, Xiangtao Liu, Yaoqin Lu and Sheng Jiang. "Association Between ABCG1/TCF7L2 and Type 2 Diabetes Mellitus: An Intervention Trial Based on a Case–Control Study." *Journal of Diabetes Research*. (February 2025).
141. Yuqin Tang, Yongqiang Zhang and Xun Hu. "Identification of Potential Hub Genes Related to Diagnosis and Prognosis of Hepatitis B Virus-Related Hepatocellular Carcinoma via Integrated Bioinformatics Analysis." *BioMed Research International*. (December 2020).
142. Yuqing Zhang, Giovanni Parmigiani and W Evan Johnson. "ComBat-seq: batch effect adjustment for RNA-seq count data." *NAR Genomics and Bioinformatics* (September 2020).
143. Zhi-Jiang Yao, Jie Dong, Yu-Jing Che, Min-Feng Zhu, Ming Wen, Ning-Ning Wang, Shan Wang, Ai-Ping Lu and Dong-Sheng Cao. "TargetNet: a web service for predicting potential drug–target interaction profiling via multi-target SAR models." *Journal of Computer-Aided Molecular Design*. (May 2016): 413 - 424.

PLAGIARISM VERIFICATION

For Title of the Thesis: "Integrating RNA-Seq analysis with machine learning approaches for the identification of hub genes affected by exposure of Bisphenol-A in patients suffering from Type 2 Diabetes"

Total Pages: 71

Name of the Scholar: Wavez Naqvi (24/BIO/15)

Supervisor: Prof. Yasha Hasija

Department: Biotechnology

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software used: Turnitin Similarity Index: 6% Total Word Count: 13084

Wavez

WM
29.05.26