

Article

Statistical Bioinformatics to Uncover the Underlying Biological Mechanisms That Linked Smoking with Type 2 Diabetes Patients Using Transcriptomic and GWAS Analysis

Abu Sayeed Md. Ripon Rouf ^{1,†}, Md. Al Amin ^{2,†} , Md. Khairul Islam ³ , Farzana Haque ⁴, Kazi Rejeev Ahmed ⁵ , Md. Ataur Rahman ^{5,6,*} , Md. Zahidul Islam ^{3,*} and Bonglee Kim ^{5,6,*} 

¹ Department of Statistics, Jagannath University, Dhaka 1100, Bangladesh; ripon1969@gmail.com

² Department of Computer Science & Engineering, Prime University, Dhaka 1216, Bangladesh; ahmedalamin2357@gmail.com

³ Department of Information & Communication Technology, Islamic University, Kushtia 7003, Bangladesh; mdkito51@gmail.com

⁴ Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Islamic University, Kushtia 7003, Bangladesh; linktofarzana@gmail.com

⁵ Department of Pathology, College of Korean Medicine, Kyung Hee University, Hoegidong Dongdaemungu, Seoul 02447, Korea; kazirejeevahmed@gmail.com

⁶ Korean Medicine-Based Drug Repositioning Cancer Research Center, College of Korean Medicine, Kyung Hee University, Seoul 02447, Korea

* Correspondence: ataur1981rahman@hotmail.com (M.A.R.); zahidimage@gmail.com (M.Z.I.); bongleekim@khu.ac.kr (B.K.)

† These authors contributed equally to this work and claimed to be combined first authors.



Citation: Ripon Rouf, A.S.M.; Amin, M.A.; Islam, M.K.; Haque, F.; Ahmed, K.R.; Rahman, M.A.; Islam, M.Z.; Kim, B. Statistical Bioinformatics to Uncover the Underlying Biological Mechanisms That Linked Smoking with Type 2 Diabetes Patients Using Transcriptomic and GWAS Analysis. *Molecules* **2022**, *27*, 4390. <https://doi.org/10.3390/molecules27144390>

Academic Editors: Valeria V. Kleandrova and Alejandro Speck-Planche

Received: 26 May 2022

Accepted: 4 July 2022

Published: 8 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Type 2 diabetes (T2D) is a chronic metabolic disease defined by insulin insensitivity corresponding to impaired insulin sensitivity, decreased insulin production, and eventually failure of beta cells in the pancreas. There is a 30–40 percent higher risk of developing T2D in active smokers. Moreover, T2D patients with active smoking may gradually develop many complications. However, there is still no significant research conducted to solve the issue. Hence, we have proposed a high-throughput network-based quantitative pipeline employing statistical methods. Transcriptomic and GWAS data were analysed and obtained from type 2 diabetes patients and active smokers. Differentially Expressed Genes (DEGs) resulted by comparing T2D patients' and smokers' tissue samples to those of healthy controls of gene expression transcriptomic datasets. We have found 55 dysregulated genes shared in people with type 2 diabetes and those who smoked, 27 of which were upregulated and 28 of which were downregulated. These identified DEGs were functionally annotated to reveal the involvement of cell-associated molecular pathways and GO terms. Moreover, protein–protein interaction analysis was conducted to discover hub proteins in the pathways. We have also identified transcriptional and post-transcriptional regulators associated with T2D and smoking. Moreover, we have analysed GWAS data and found 57 common biomarker genes between T2D and smokers. Then, Transcriptomic and GWAS analyses are compared for more robust outcomes and identified 1 significant common gene, 19 shared significant pathways and 12 shared significant GOs. Finally, we have discovered protein–drug interactions for our identified biomarkers.

Keywords: Type-2 diabetes; smoking; bioinformatics; association; GSEA; pathway; orthology

1. Introduction

Type 2 diabetes (T2D) is a chronic metabolic disease defined by insulin insensitivity corresponding to impaired insulin sensitivity, decreased insulin production, and eventually, failed beta cells in the pancreas [1,2]. So, T2D occurs when the body cannot effectively utilise insulin [3]. According to current forecasts, the total number of diabetes patients would be

greater than 50% from 2017 to 2045, resulting in over 693 million or 0.693 billion people with diabetes. Therefore, the health expenditure will be about 850 billion US dollar [4]. It is unpleasantly surprising that in diabetes patients, more than 95%, have T2D [5]. In consequence, diabetes was responsible for about five million deaths in 2017 among people aged 20 to 99 years over the world [4].

It is clinically statistically shown that professional smokers are more likely to develop T2D with a 30–40% higher chance than nonsmokers [6]. Another research study demonstrated the smoking prevalence in China by examining 1658 people, including 621 (37.5%) non-smokers and 1037 (62.5%) people actively smoking. Whereas diabetes was found to be more related to active smoker than the non-smoker [7]. Not only that, both men and women smokers were much more likely to establish T2D than those who never smoked [8,9]. Moreover, diabetes patients with a smoking history have increased risk of developing complications such as heart and kidney disease, and nerve damage, as well as alleviated blood supply to the legs and feet [6]. The reason for the relation between smoking and T2D is the dysfunction of beta-cells, where smoking may harm the act of beta-cells [10,11]. Similarly, it is also visible that the beta-cells are harmed in T2D patients [2]. Due to the damage of beta-cells, there is an association between the increase of T2D and insulin resistance, which can be altered by smoking in both a direct and an indirect way [12,13]. According to a new theory, Cigarette smoking was associated with lower vitamin D levels [14,15], where 83.2% of patients with T2D have vitamin D deficiency, which is a significant number [16]. Another represented that a higher casualty of T2D is connected with lower HDL cholesterol levels [17,18], and smoking cigarettes is also linked to lower HDL cholesterol levels [19,20]. In addition, increased (triglycerides) TG levels were found to be substantially linked to the development of T2D [21,22], whether smoking helps increases triglycerides [23].

The development of T2D can be amplified by regular smoking, as well as many complications which can be developed in T2D patients with a smoking history. Hence, we have tried to retrieve common molecular mechanisms that exist between T2D and smokers. Molecular mechanisms and relationships underlying the T2D-smoker are yet unclear in medical endocrinology [24], but they are still of significant interest. Moreover, there is still a lack of bioinformatics studies regarding the relationship between T2D and smoking. The goal of our study is to warn the active smokers that they can develop T2D and to help the drug developers so they will be able to eliminate the risk of many complications in T2D patients due to smoking. Therefore, this inspired us to apply a bioinformatics-based systemic pipeline to analyse the gene expression data of T2D patients and smokers to obtain an insightful understanding of their relationship. To comprehend molecular causes of biological disease and condition, differentially expressed genes (DEGs) [25] should be identified for further system biology analysis. In this study, RNA-seq datasets [26] (transcriptomic data [27]) were utilised to discover DEGs shared by T2D and smoker. We have used the DESeq2 package [28] which is frequently used for differential gene expression analysis of RNA-seq count data. We identified commonly dysregulated genes, pathways and Gene Ontology (GO) protein–protein interaction (PPI) [28], hub-proteins [29], Transcription Factor Gene Interactions [30], Gene miRNA Interaction [31], Protein–Drug Interaction [32]; that are associated with T2D and smoker using a systems biology approach. Finally, we have analysed common genes, pathways and GO using GWAS datasets [33]; later, we discuss and compare with transcriptomic analysis. We have represented the methodology of our hypothesis as shown Figure 1.

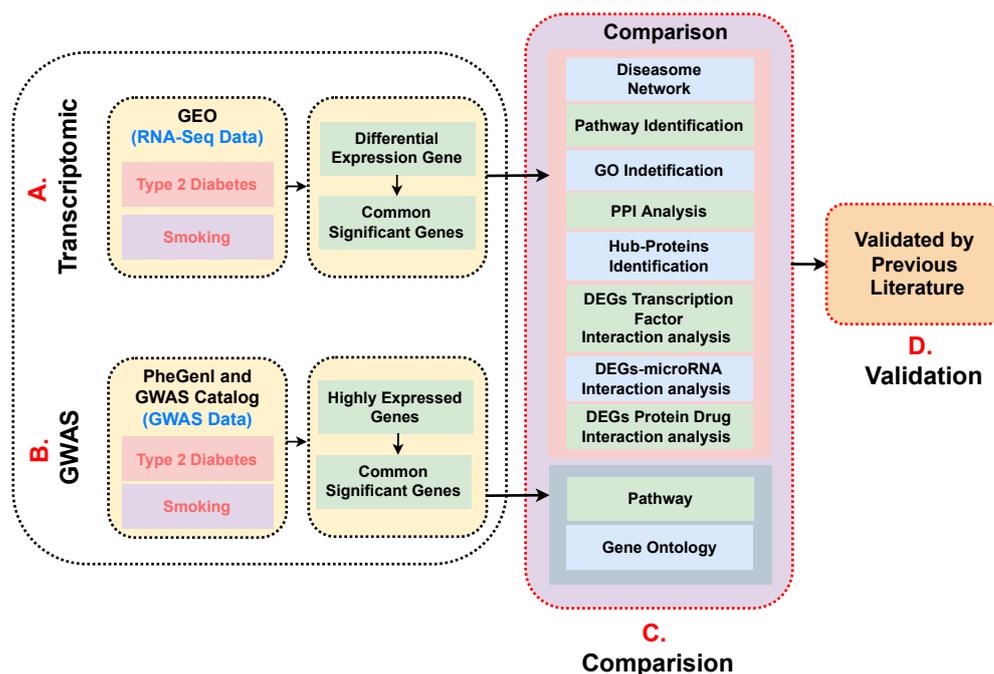


Figure 1. The figure demonstrate the working flow of our proposed methodology. **A.** Collected Transcriptomic datasets from GEO for Type 2 diabetes and Smoking. Generate DEGs and find out shared significant genes between both conditions **B.** Collected GWAS datasets from PheGenI and GWAS Catalog databases for Type 2 diabetes and Smoking. Generate DEGs and find out common significant genes between the conditions. **C.** Statistical analyses on Transcriptomic data to identify Pathway, Gene Ontology, Disease Network, PPI Analysis, Hub-Proteins, DEGs Transcription Factor Interaction, DEGs–microRNA Interaction, and Protein–Drug Interaction. Comparison between Transcriptomic and GWAS analysis based on Pathway and Gene Ontology. **D.** Validated the results by previous literature to find the biological relevance.

2. Materials and Methods

2.1. Datasets Employed in This Study

We used data from the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) [34]. Datasets for our work are also available in GREIN, a web-based interactive platform that offers simple substitutes for exploring and analysing GEO RNA-seq data [26]. GREIN was utilized in many more studies to perform differential expression analysis of their RNA-seq dataset [35]. Many datasets were returned for each disease while searching in NCBI; however, most of them were rejected since they did not fulfil the sample size where the size should have been 6; similarly, neither replicate datasets or RNAseq datasets which were not taken from homo sapien organisms, as well as datasets, do not include both a Case and Control. In six samples, there should be at least three case and three control. For this type of study, we looked for datasets with the lowest amount of bias and noise. Here, after many searches, we have selected datasets that overcome any noise as much as possible for this analysis. This procedure selected two datasets that are suitable for our investigation and are very relevant to T2D and smoking.

Accession numbers of the datasets, which is human gene expression, are GSE106177 [36] and GSE47718 [37]. T2D (GSE106177) dataset is collected from the DNA of human primary cardiac mesenchymal cells (CMSC) from seven diabetic (D) donors and seven non-diabetic (ND) donors were analyzed. The genomic DNA of human CMSCs isolated from diabetes donors had an accumulation of 5-methylcytosine, 5-hydroxymethylcytosine, and 5-formylcytosine, as determined by quantitative global analysis, methylated and hydroxymethylated DNA sequencing, and gene-specific GC methylation detection. The smoking datasets (GSE47718) are generated from the human airway basal cell (BC) transcriptome of seven smokers and

seven non-smokers. On 19q13.2 (CYP2F1 and RASGRP4), two airway epithelium genes were also identified to study hypermethylation in smokers in comparison with non regular smokers (shown in Table 1). As well as this, BC of the case and control were compared utilizing massive parallel RNA sequencing.

Table 1. An overview of the data and findings from the transcriptome study. It includes the dataset accession number, sample source, total raw genes, sample number, significant genes.

Disease Name	GEO Platform	Tissues/Cells	GEO Accession	RAW Genes	Case Samples	Control Samples	Significant	Up Reg. Genes	Down Reg. Genes
Type-2 Diabetes (T2D)	Illumina NextSeq 500 (Homo sapiens)	Human cardiac mesenchymal cells	GSE-106177	18,619	7	7	1367	768	599
Smoking	Illumina HiSeq 2000 (Homo sapiens)	Airway Basal Cells	GSE-47718	21,178	7	10	962	682	280

2.2. Preprocessing of Raw Counts and Its Differentially Expression Analysis

The GREIN Gene Expression Omnibus(GEO) provided us with gene expression RNA-sequence datasets. The DESeq2, an R package, was used to identify differentially expressed genes (DEGs) of T2D and Smoking related RNA-Seq count data [28]. In DESeq2, the geometric mean of every single gene was calculated among all samples to normalize the data. Then, DESeq2 used the cook's distance to automatically filter out lowly expressed and outlying genes. Based on two conditions such as $\|log_2FC\| \geq 1$ and adjusted p -value ≤ 0.05 , the significant DEGs were filtered. By utilising the match function, we were able to find shared DEGs between T2D and smoking and create a Venn diagram (Shown in Figure 2).

2.3. Identification of Molecular Pathway and Gene Ontology

Many bioinformatic techniques often focus on measuring the relevance of gene set similarity with previously annotated gene sets in order to find functional pathways connected to a particular disease [38]. We used gene set enrichment analysis with EnrichR to find out its associated pathways and Gene Ontology (GO) terminologies by using the overlaying DEGs between T2D and smokers. The findings identified pathways and GOs, which would definitely help us to gain an appropriate understanding of the biological mechanism related to both conditions [39]. For biological domains, the underlying purpose of the GO project is to develop a structured gene dictionary that may be used to explain gene-products in any living life form. In terms of GO terminology, there are three categories: biological process (BP), cellular component (CC), and molecular function (MF) [40]; however, we only considered the BP for our study. Pathways demonstrate a crucial role in how organisms respond to stimuli. In life science, pathway analysis is commonly used to help researchers understanding the high-throughput biological data through their underlying molecular mechanisms. It may also describe the relation of diseases or conditions to each other [41].

2.4. GWAS Data of Type 2 Diabetes Mined to Compare with the GWAS Data of Smoker

Researchers use genome-wide association studies (GWAS) to examine a group of genetic mutants which are linked with a certain illness or condition. During this procedure, the genomic sequences of a large number of patients were collected and evaluated so that we could determine the DNA polymorphisms that are present in the genomes of the patients. We have collected GWAS data from both T2D patients and smokers. We have collected the GWAS data from two well-known databases such as PheGenI [42] and GWAS Catalog [43]. Both the databases created using previous studies that incorporated GWAS results published earlier. We have only considered data for T2D patients and smokers. Moreover, we have filtered the most significant data by a threshold p -value that is less than

1.0×10^{-8} . We also identified the GOs and pathways associated with the identified common significant genes from GWAS analysis. Finally, we have compared the transcriptomic analysis result with the GWAS analysis to better understand and relate the common biological outcomes by both studies.

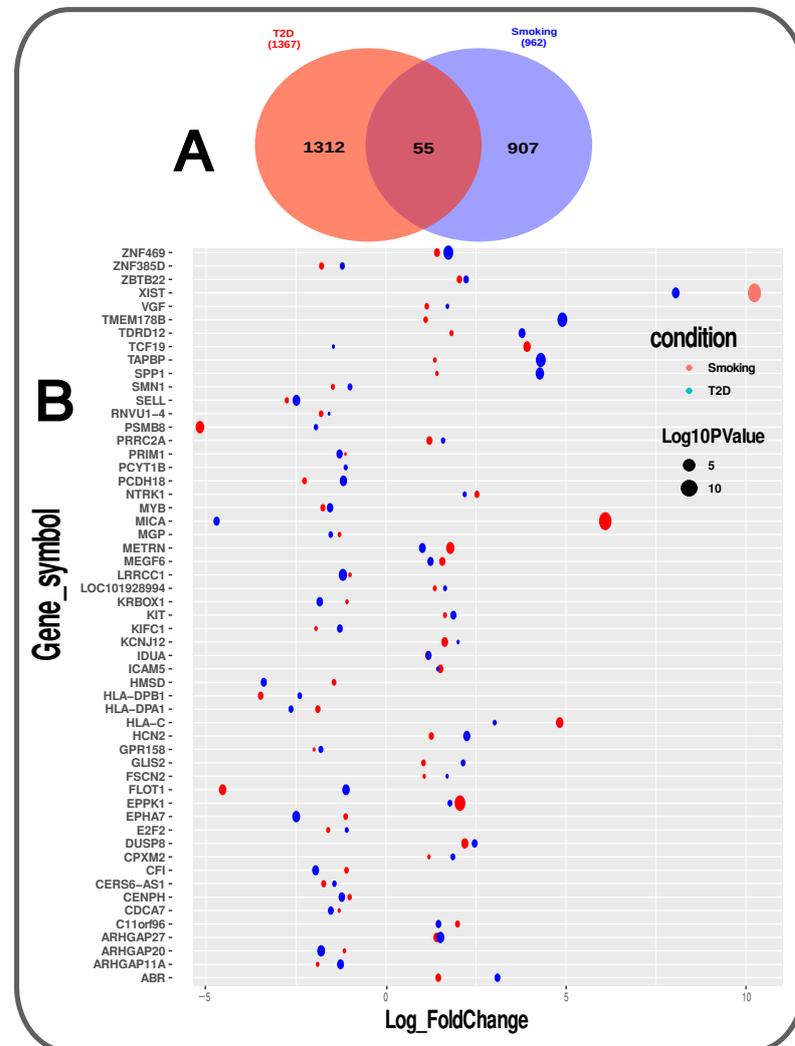


Figure 2. The figure represented the summary of transcriptomic analysis. **A.** Common biomarker Genes between T2D and Smoking using a venn diagram **B.** Bubble plot of common biomarker genes and their associated adjusted p -value and Log2 fold change.

3. Results

3.1. Gene Expression Analysis of Transcriptomic Data

We have collected RNA-Seq data from Grein or NCBI to examine the impact of gene expression on T2D patients and Smokers. As shown in Table 1, “Illumina Next Seq 500 (Homo sapiens)” GEO platform has provided the T2D data, and the “Illumina HiSeq 2000 (Homo sapiens)” has provided the smoking data. The T2D RNA-seq data have been collected from “Human cardiac mesenchymal cells”, and the smoking RNA-seq data have been collected from “Airway Basal Cells”. The GEO accession ID-GSE106177 is selected for T2D [36], and GEO accession ID-GSE47718 is selected for Smoking [37]. In T2D, 14 samples were included, seven for the case and seven for control. The total differentially expressed (DE) genes for T2D is 18,619, which is found after differential expression analysis; which is also called generated signature data. Then, two conditions considered such as “Adjusted PValue” and “Log2 Fold-Change” on signature data. We have found 1367 significant genes after applying the condition whereas Adjusted PValue is less than 0.05 and $\text{abs}(\text{LogFC})$

is greater than or equal 1.0. Among significant genes, 768 is Up-regulated genes, and 599 is downregulated genes. On the other hand, a total of 17 samples are included in the Smoking dataset, whereas 7 for the case (regular-smokers) and 10 for control (non-smokers). After generating signature data, we have found 21178 DE genes. Then applied the same conditions that applied to the T2D, and we have found 962 significant genes. Among these, 682 is positively expressed (Up-regulated), and 280 is negatively expressed (downregulated).

Then, we have compared the upregulated genes of T2D with the upregulated genes of smokers. Similarly, we have compared the downregulated genes of both conditions. Between T2D and Smoking, the number of upregulated common genes are 27 and the number of downregulated common genes are 28 (Shown in Figure 3). The most significant shared up regulated genes are *TMEM178B*, *TAPBP*, *ZNF469*, *SPP1*, *ARHGAP27*, *XIST*, *HCN2*, *TDRD12*, *METRIN*, *MEGF6*, *KIT*, *IDUA*, *DUSP8*, *ABR*, *C11orf96*, *ZBTB22*, *GLIS2*, *EPPK1*, *CPXM2*, *PRRC2A*, *NTRK1*, *HLA-C*, *LOC101928994*, *VGF*, *ICAM5*, *FSCN2* and *KCNJ12*. Moreover, the most significant shared down regulated genes between T2D and Smoking: *LRRCC1*, *EPHA7*, *ARHGAP20*, *SELL*, *FLOT1*, *PCDH18*, *CFI*, *ARHGAP11A*, *CENPH*, *KRBOX1*, *MYB*, *PRIM1*, *MICA*, *HMSD*, *CDCA7*, *KIFC1*, *HLA-DPA1*, *ZNF385D*, *GPR158*, *SMN1*, *HLA-DPB1*, *MGP*, *CERS6-AS1*, *PSMB8*, *PCYT1B*, *E2F2*, *TCF19* and *RNVU1-4*.

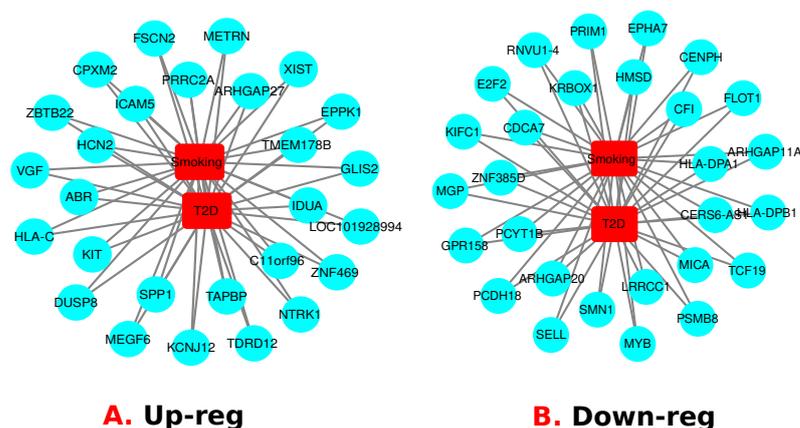


Figure 3. Upregulated and downregulated genes between T2D patients and smokers are represented separately.

3.2. Pathway and GO Related Functional Association Analysis

In pathway analysis, we have used five different databases named BioCarta, Elsevier Pathway, KEEG, Reactome and WikiPathway. Then, we have looked for highly expressed pathways by utilizing shared DEGs between T2D and smoking. These pathways were then divided into functional groups for analysis. Initially, we got 573 signalling pathways linked with both disease and condition. Manual curation was then utilised to abate the number of pathways. Pathways with p -values below 0.05 are considered. Therefore, we got 169 most significant signalling pathways. Finally, we sorted the pathways in ascending order depending on p -value. The top 20 pathways were represented in Figure 4 that were linked to T2D and smoking condition.

The GO approach uses the biological process (BP), cellular component (CC), and molecular function (MF) to categorise them into functional categories, but we only consider the BP database of GO terminologies. At first, 489 GO terms found common between T2d patients and smokers. After that, the terms with a p -value less than 0.05 were included as the most significant GO terms. A total of 137 GO terminologies are found as the mostly enriched GO terms between the conditions. Figure 5 summarises the top 20 most significant GO terms of BP between T2D and smoking.

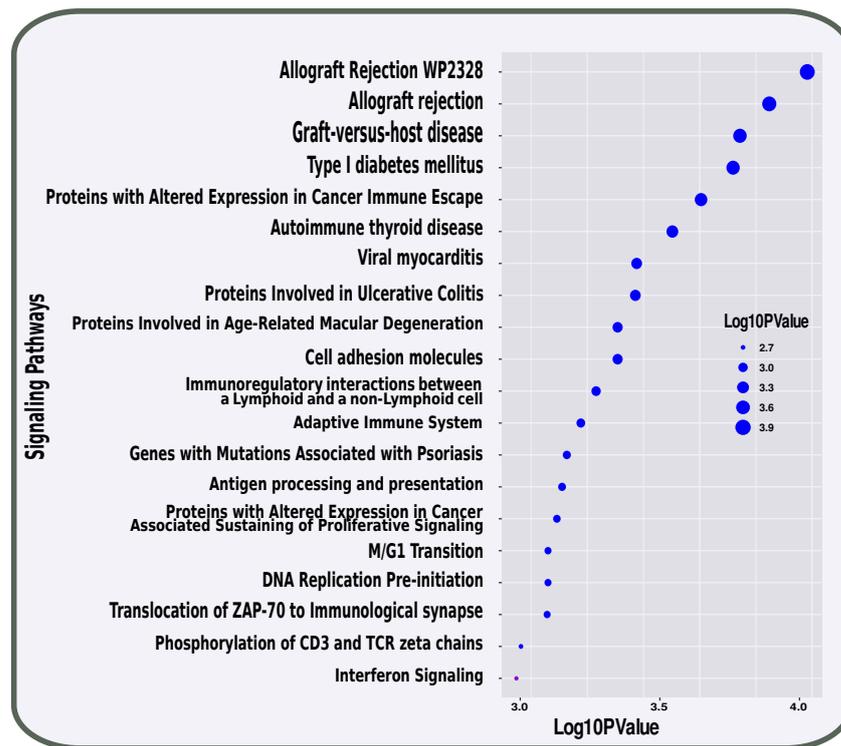


Figure 4. Top 20 pathways are represented using a bubble plot that are associated with both conditions in transcriptomic analysis.

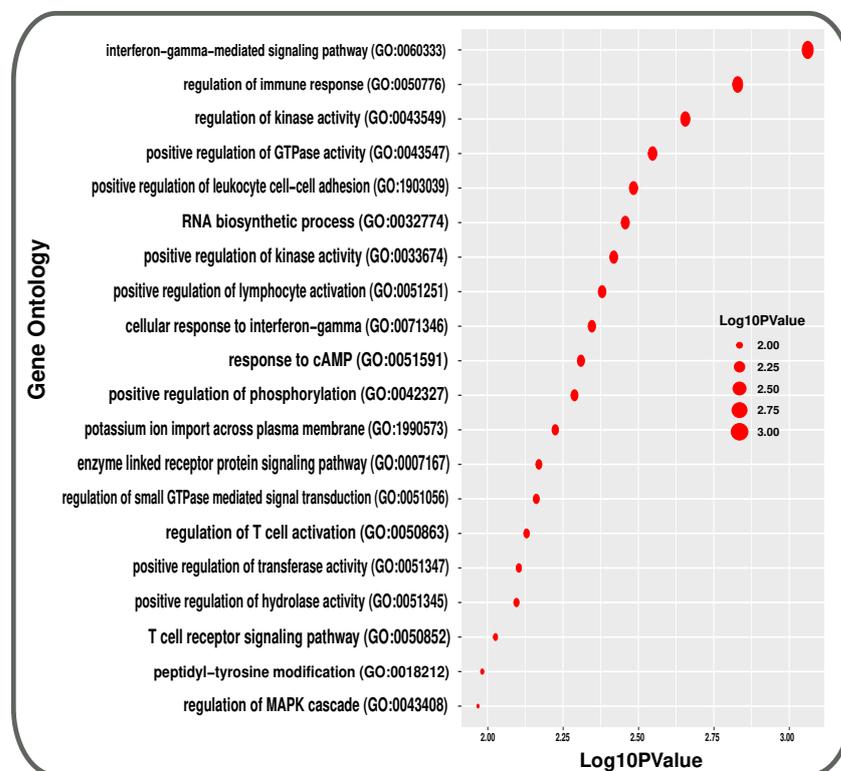


Figure 5. Top 20 Gene Ontologies represented using a bubble plot that are linked with both conditions in transcriptomic analysis.

3.3. Protein-Protein Interactions (PPIs) Analysis

STRING [44], a database, was handled to perform protein-protein interactions (PPI). In Cytoscape [45], the PPI network has been processed and analysed. Based on proteins

encoded by the common DEGs between T2D and Smoking, we developed a PPI network on the STRING. In Cytoscape, this is prepared and assessed, and the results are shown in Figure 6. Protein subnetworks that are shared by two or more diseases are known to be linked to each other. The highly interacting proteins were discovered by PPI analysis using topological characteristics, such as degree larger than 15°. The CytoHubba [46], a plugin in Cytoscape, was installed and operate to find the highly linked hub proteins in the PPI network using degree and Maximal CliqueCentrality (MCC) and BottleNeck algorithm. These newly discovered hub proteins may be beneficial as therapeutic targets, but further study is needed to understand their functions. As mentioned earlier, hub proteins are identified using two algorithm named MCC and BottleNeck, and represented in Figure 7A,B. By using the MCC algorithm we have found a total of 23 hub proteins where the top 10 genes are highly significant marked as red, orange and yellow colors. In addition, by using the BottleNeck algorithm, we have found a total of 20 hub proteins where 10 are shown as the most significant (marked as yellow, red and orange colors). In both algorithms, we have found a total of 13 most significant hub proteins that are unique and they are: KIFC1, TAPBP, HLA-DPA1, HLA-DPB1, PSMB8, SPP1, KIT, MYB, MICA, PRIM1, TCF19, HLA-C and ZBTB22. Among them seven proteins are shared by both algorithms that are: KIFC1, TAPBP, HLA-C, HLA-DPB1, HLA-DPA1, KIT and PRIM1 as well as most significant.

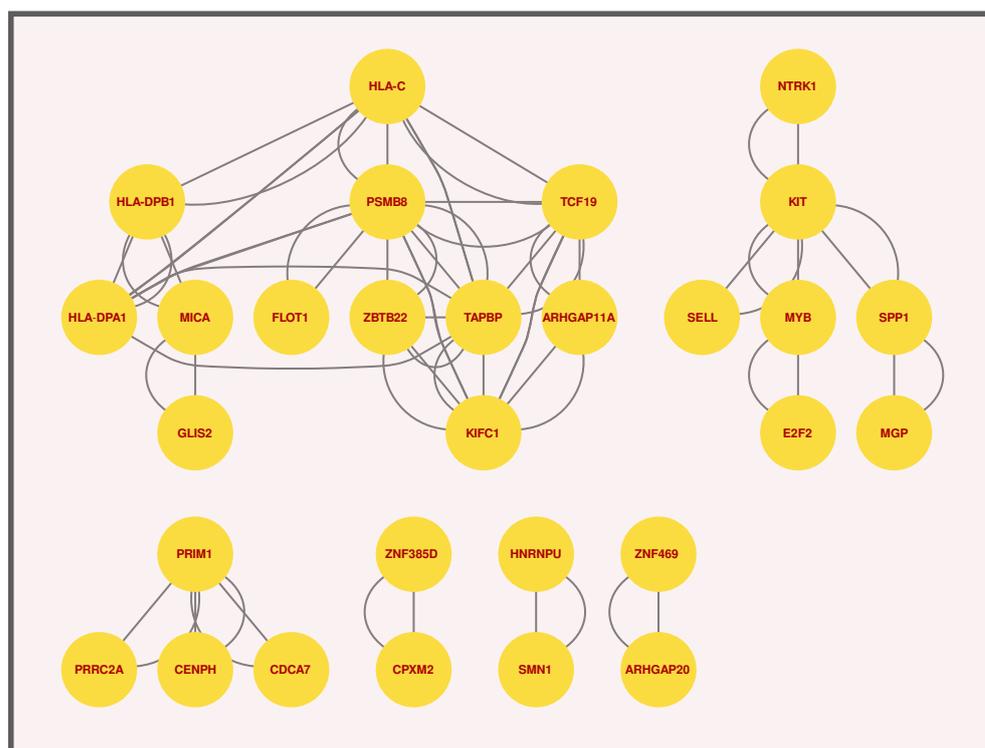


Figure 6. The figure illustrate the Protein–Protein Interactions between T2D and Smoking.

3.4. Identification of Transcriptional and Post-Transcriptional Regulators of the Differentially Expressed Genes

For TFs, protein–drug interactions, and gene–microRNA interactions, we have used Net-workAnalyst tools [47]. Transcription factors (TFs) are proteins that control gene expression in all living organisms. In every cell process, TFs perform a critical function [48]. Gene expression is regulated in the post-transcription stage by non-coding short RNA molecules called miRNAs. Protein–drug interaction studies are critical to understanding the structural characteristics required for ligand affinity [49,50]. The TFs–Gene network was constructed using the ChIP-X [51] and JASPAR database [52]. The TarBase [53] and miRTarBase database [54] were utilized to develop Gene–miRNAs interaction network

using NetworkAnalyst. Moreover, protein–drug interactions are constructed using the DrugBank database [55]. Figure 8 visually represents the TF–Gene Interactions, and the regulator genes are: *ARHGAP27*, *E2F2*, *KIT*, *EPHA7*, *TAPBP*, *KIFC1*, *ARHGAP11A*, *SPI1*, *POU5F1*, *NANOG*, *SOX2*, *METRN*, *TCF19*, *SMN1*, *TDRD12*, *MGP*, *E2F1*, *GATA2*, *NFKB1*, *NFIC*, *SRF*, *YY1*, *TEAP2A*, *FOXL1*, *TP53*, *POU2F2*, *FOXC1*, *PRRC2A*, *VGF*, *CDCA7*, *SPP1*, *FLOT1*, *ABR* and *MYB*. Figure 9 visually represents miRNA’s genes interactions that are: mir-146a- 5p, mir-34a-5p, mir-27a-3p, mir-129-2-3p, mir-124-3p, mir-193b-3p, let-7b- 5p, mir-26b-5p, mir-31-5p, mir-93-5p, mir-331-3p. Finally, Figure 10 visually represents the Protein–Drug Interactions for smoker and T2D complications and common genes are: *KIT* and *NTRK1*. The genes are related to drugs called Ponatinib, Dasatinib, Pazopanib, Sorafenib, ABT-869, OSI-930, MP470, XL184, XL820, Nilotinib, Phosphotyrosine, Lenvatinib, Sunitinib, Imatinib, Regorafenib and Amitriptyline.

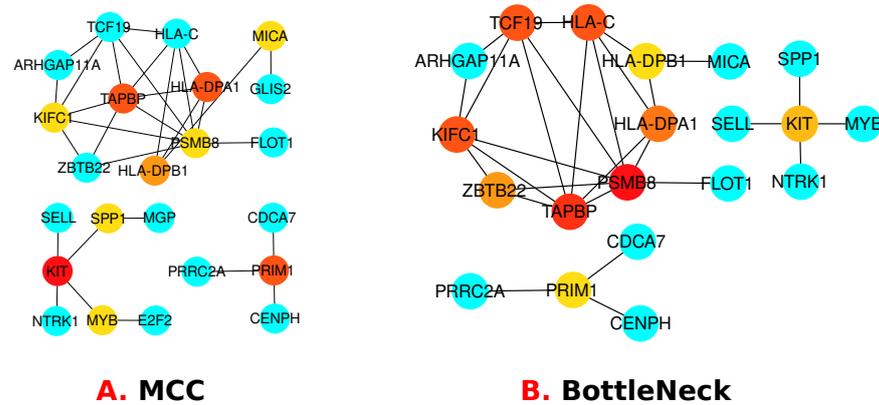


Figure 7. Hub proteins identified using 2 different cyto-hubba algorithms called MCC and BottleNeck to demonstrate the association between T2D and Smoking.

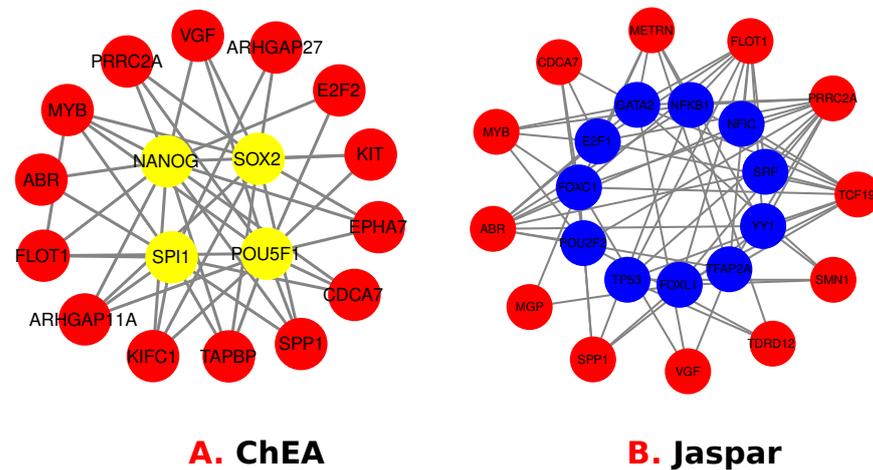


Figure 8. TF–Gene interactions showed using 2 different algorithms called ChEA and Jasp to illustrate the linking between T2D and Smoking.

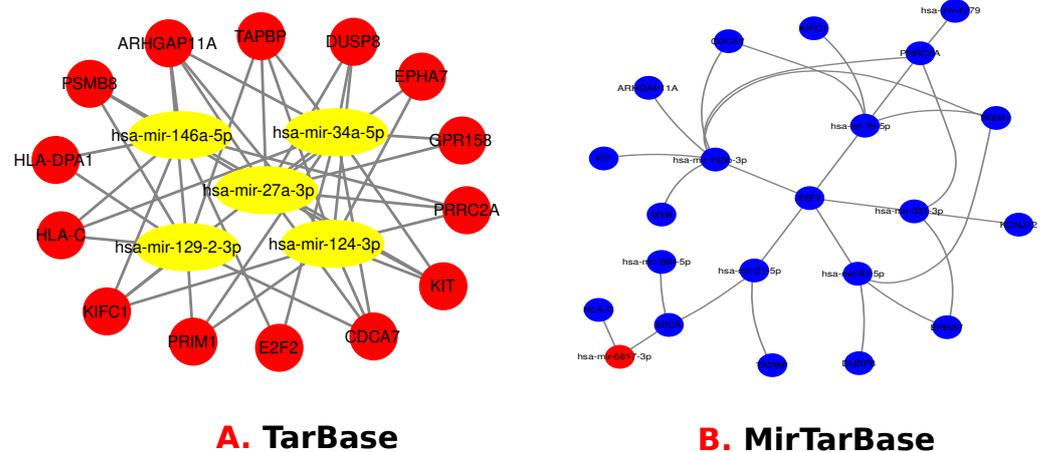


Figure 9. Gene miRNA is identified between T2D and Smoking using 2 different algorithm called TarBase and MirTarBase.

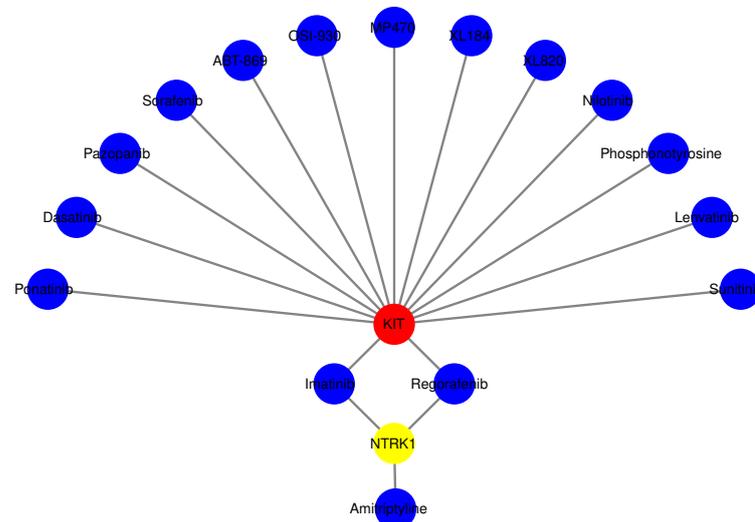


Figure 10. The figure shows the Protein–Drug Interaction between T2D and smoking using Protein–Drug Interaction algorithm.

3.5. GWAS Analysis of Type 2 Diabetes with Smoker and Comparison with Transcriptomic Analysis

GWAS analysis was performed to find the most significant genes for both T2D and smokers. We employed a threshold p -value which is less than 1.0×10^{-8} . After that, we found 595 significant genes for T2D and 535 significant genes for smokers. Then, we compared both significant genes to obtain the candidate genes associated with both conditions. There are 57 candidate genes (Shown in Figure 11) that are shared by both conditions: *GNPDA2*, *ITPR2*, *JAZF1*, *MSRA*, *ZC3H4*, *NRXN3*, *MAP2K5*, *TCF7L2*, *HSD17B12*, *CCND2*, *AUTS2*, *FTO*, *MC4R*, *ANKRD55*, *PPARG*, *TMEM18*, *VEGFA*, *TFAP2B*, *CTTNBP2*, *PINX1*, *C5orf67*, *LYPLAL1*, *MIR4432HG*, *COBLL1*, *ADAMTS9-AS2*, *NFAT5*, *CMIP*, *ALDH2*, *ZFH3*, *SEC16B*, *TSEN15*, *SIX3*, *BCL11A*, *ZBTB20*, *FGFR4*, *HMGA1*, *CALCR*, *BDNF*, *FAIM2*, *TRIB1*, *ARID5B*, *ADAMTS9*, *LINGO2*, *EHMT2*, *ZBTB38*, *THAP12P9*, *Metazoa_SRP*, *CCND2-AS1*, *HLA-C*, *SPPL3*, *LINC02537*, *RFLNA*, *TSEN2*, *ZC3H11B*, *RSPO3*, *NYAP2* and *CUX2*. Moreover, we have used the identified biomarker genes from GWAS analysis in order to find the signalling pathways and GO terms. We have used the same tools as transcription profiles for these analysis. In GWAS analysis, we have got 839 signalling pathways and 802 GOs shared by both T2D and smoker. Moreover, Figures 12 and 13 represent the top 20 GOs and pathways respectively among them. We have also used the same condition

(p -value ≤ 0.05) to distinguish the most significant pathways and GOs. Thus, the number of pathways reduced to 196; similarly, GOs reduced to 300. A comparison performed between transcriptomic and GWAS analysis where we have found only one biomarker gene (HLA-C) common between two studies. We have also compared significant pathways from both transcriptomic and GWAS analysis and found 19 significant pathways are shared by both studies: Genes with Mutations Associated with Psoriasis, Proteins Involved in Ulcerative Colitis, Proteins Involved in Psoriatic Arthritis, Proteins Involved in Neuroblastoma, Proteins Involved in Spontaneous Abortion, Proteins with Altered Expression in Osteoarthritis, NTRK FOXO/MYCN Signalling, NTRK1/2/3 Acetylcholine Production, Receptors and Adaptor Proteins Activated in Cancer, Proteins Involved in Glioma, Kaposi sarcoma-associated herpesvirus infection, MAPK signalling pathway, PI3K-Akt signalling pathway, Endosomal/Vacuolar pathway, VEGFR2 mediated cell proliferation, PI3K-Akt signalling pathway WP4172, Allograft Rejection WP2328, Pathways Regulating Hippo Signalling WP4540 and NOTCH1 regulation of endothelial cell calcification WP3413. Similarly, 12 significant GOs are shared by both studies: regulation of neuron apoptotic process (GO:0043523), positive regulation of phosphorylation (GO:0042327), antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway (GO:0002484), nerve growth factor signalling pathway (GO:0038180), negative regulation of sprouting angiogenesis (GO:1903671), positive regulation of bone resorption (GO:0045780), positive regulation of DNA-binding transcription factor activity (GO:0051091), ovarian follicle development (GO:0001541), neurotrophin TRK receptor signalling pathway (GO:0048011), positive regulation of histone H3-K4 methylation (GO:0051571), positive regulation of cell projection organization (GO:0031346) and sympathetic nervous system development (GO:0048485).

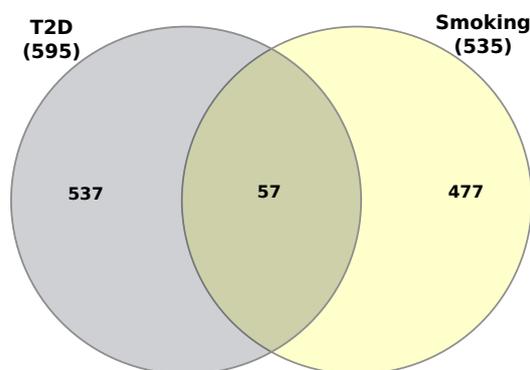


Figure 11. The Venn Diagram represented the shared candidate/biomarker genes found between Type 2 diabetes and smokers from GWAS studies.

3.6. Validating Potential Biomarker Targets Using Earlier Literatures

We investigated the previously published literature to verify the presence of biomarker genes that are found both in smokers and T2D (Shown in Table 2). The expression of SPP1 can be boosted by smoking cigarettes [56] and SPP1 expression similarly raised in patients with T2D [57]. Smoking has been linked to MICA [58]. PSMB8 was found linked with regular smokers [59]. The protein PSMB8 was observed in the proteasome that is responsible for ATP-dependent protein degradation, which is a direct reason for developing T2D [60]. Patients with T2D had a higher level of XIST expression than those without diabetes [61]. Type 2 diabetes risk is substantially increased by DUSP8 expression in individuals [62]. Patients with T2D are more likely to experience ABR abnormalities than those without the disease [63]. In both type 2 diabetes and atherosclerosis, the protein PRRC2A was found to be present [64]. Compared to non-smokers, heavy smokers show a substantial decrease in the transcription of PRRC2A genes [65]. The NTRK1 has been identified in T2D [66]. It is possible that the EPHA7 gene is linked to smoking [67]. T2DM may be impacted by SELL [68]. T2D patients are linked with PRIM1 deficiency [69]. Carcinogenic pathways

caused by PRIM1 may be influenced by factors in the tumor microenvironment, such as smoking (nicotine) [70]. There are genes called KIFC1 that are connected with a person's smoking status [71]. ZNF385D exhibited increased levels of methylation in people who are currently smoking [72]. In T2DM, the MHC-class II molecules, particularly HLA-DPB1, which encode human leukocyte antigens, were dramatically increased [73].

Table 2. Identified possible target genes or biomarkers that are shared by smoking and type 2 diabetes have been confirmed by previous studies.

Gene	Smoking	T2D
SPP1	[56]	[57]
MICA	[58]	-
PSMB8	[59]	[60]
XIST	-	[61]
DUSP8	-	[62]
ABR	-	[63]
PRRC2A	[65]	[64]
NTRK1	-	[66]
EPHA7	[67]	-
SELL	-	[68]
PRIM1	[70]	[69]
KIFC1	[71]	-
ZNF385D	[72]	-
HLA-DPB1	-	[73]

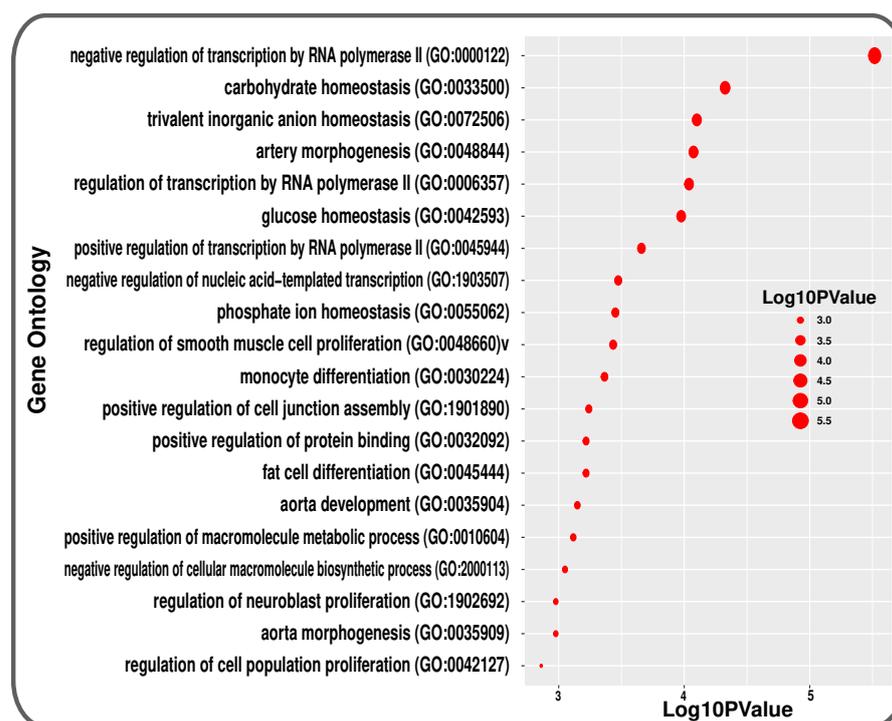


Figure 12. Top 20 GOs are represented using a bubble plot that are associated with both conditions in GWAS analysis.

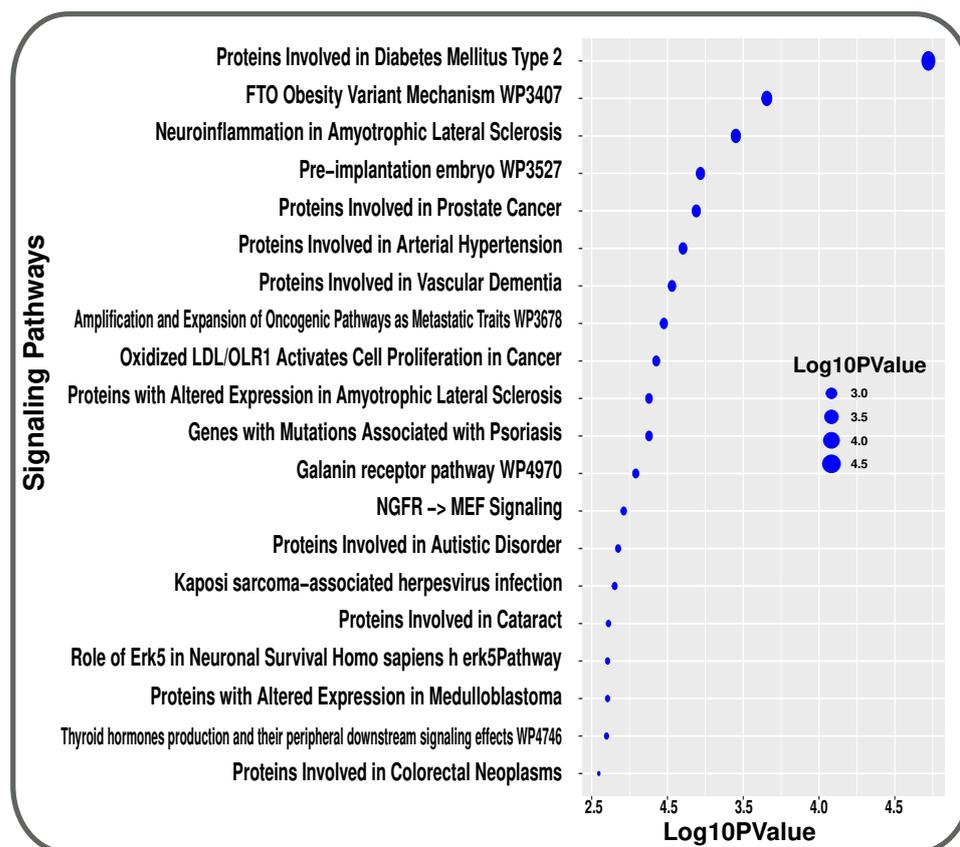


Figure 13. Top 20 pathways are represented using a bubble plot that are associated with both conditions in GWAS analysis.

4. Discussion

There is no doubt that T2D is a major threat to humanity. People who smoke cigarettes have a 30–40% increased chance of acquiring T2D in comparison with nonsmokers [6]. Moreover, we have highlighted through a study that many complications can be developed such as kidney and heart disease in T2D with smoking addiction [6]. So, we can say that it is possible that active smokers may easily develop T2D and it is also possible that T2D patients with a smoking history may develop many complications or diseases. Therefore, our study aims to extract genetic correlations between smoking and T2D. Furthermore, our predicted drugs might be the potential therapeutics for T2D patients with a smoking history. Not only that, but also physicians would inspire the patients to stay away from smoking. This would definitely eliminate the risk of developing T2D in individuals or other complications caused in T2D patients. The bioinformatics approach provides a comprehensive understanding of the molecular mechanisms in disease progression. In this study, the investigation was carried out on both transcriptomic and GWAS profiles of smokers and T2D.

To investigate if there was any significant dysregulation, we have performed Differential Expression Analysis (DEA) followed by identification of common genes including up or downregulated genes (shown in Figures 2 and 3, respectively), gene ontology (GO) (Shown in Figure 5), pathways (represented in Figure 4), protein–protein interactions (shown in Figure 6), hub–protein interactions (shown in Figure 7), transcription factor gene interactions (represented in Figure 8), gene–miRNA interactions (shown in Figure 9) and protein–drug interactions (represented in Figure 10) using transcriptomic profiles of smoking and T2D. As well as, we have identified common genes, GOs and pathways using GWAS profiles of smoking and T2D (respectively, shown in Figures 11–13). GWAS data analysis provides a more robust understanding of our hypothesis as well as compared with transcriptomic profiles. The T2D transcriptomic dataset was collected from T2D patients

and non-diabetic healthy individuals, and the smoking dataset was collected from smokers and nonsmokers (as shown in Table 1). We also verified our results by previous literature published in various journals. The flow diagram of our methodology has been visually represented and outlined with proper direction in Figure 1.

First of all, we have focused on four pathways, namely PI3K/AKT signalling pathway, MAPK signalling pathway, Endosomal/Vacuolar pathways and Allograft Rejection pathway, which are found by the shared profile of T2D and smoker. These pathways were resulted from both transcriptomic and GWAS analysis.

T2D development is significantly influenced by the MAPK signalling pathway [74], where cigarette smoke particles may activate MAPK pathways [75]. The Allograft Rejection pathway is also activated in T2D patients [76], where Allograft Rejection is more likely to active in people with a history of smoking cigarettes [77,78]. PI3K/AKT pathway damaged in various tissues of the body leads to the development of T2D as the result of insulin resistance [79], and tobacco smoke exposure can activate PI3K/Akt pathway [80]. Endosomal/Vacuolar pathways is presented in T2D [81]. Hence, in vitro, researchers can work on this pathway to make resistance to T2D. Therefore, the continuous inactive stage of the Allograft Rejection pathways, PI3K/Akt pathway and MAPK pathway may be possible by quitting smoking at an early period or taking drugs to alter the current active status of the pathways in T2D patients. Ultimately, avoiding smoking is highly recommended for both control (absence of T2D) and T2D patients so that the pathways do not influence by smoking. As far as we know, no bioinformatics approach had previously described these pathways and to the best of my knowledge, other pathways identified in our study are not specified by any previous literature for developing T2D. Yet those pathways may be potential drug targets that paved the way for further research and analysis.

For significant common genes, we have focused on certain genes named SPP1, PSMB8, PRRC2A, and PRIM1 that are found from transcriptomic analysis. These genes are highly connected with each other, which also resulted in hub-protein analysis (as shown in Figure 7). These genes have been picked up based on the previously published literature related to both smokers and T2D patients (Shown in Table 2). SPP1 has been linked to an increased risk of long-term effects of diabetes and has been found to be elevated in T2D patients [57]. Where, SPP1's expression in induced sputum can be raised by smoking cigarettes, and the amount of SPP1 is also increased in induced sputum [56]. PSMB8 was found to be associated with insulin-dependent Diabetic Mellitus [82]. Moreover, it is highly expressed among smokers [59]. The PRRC2A gene was responsible for the development of insulin-dependent diabetes mellitus [83]. It was found that nicotine therapy increased the expression of PRIM1 protein [70]. Moreover, PRIM1 mutation has also been recorded for developing T2D [69]. As the specified genes such as PRIM1, SPP1, PSMB8 and PRRC2A are activated in T2D patients, it is essential to remove the mutation of these genes to rehabilitate. However, it is smoking that regulates the activation of these genes in individuals as well. Thus, the effect of smoking is a serious issue for one's health and needs to avoid in the first place. As far as we know, these genes were not analysed previously in any bioinformatics study.

In protein-drug interactions, we have discovered two genes named KIT and NTRK1 that are connected to 16 drugs for T2D with a smoking history (shown in Figure 10). KDM6A binding to the NTRK1 promoter because of YY1 and subsequent TRKA overexpression led to imatinib resistance [84]. Moreover, it has been demonstrated that imatinib improves glycemic control in diabetics [85]. According to this information, imatinib is connected to the NTRK1 gene, and it helps to control T2D. Comorbid diabetes mellitus, IFG, or IGT can be treated with sorafenib, a safe and effective therapeutic option [86]. In addition, sorafenib inhibits the Imatinib-Resistant KIT. As well, stable Ba/F3 KIT mutants recapitulated the genotype of imatinib-resistant individuals with primary and secondary KIT mutations [87]. So, this sorafenib drug can resist T2D. It was found that sunitinib lowered insulin requirements [88]. In addition, tyrosine kinase inhibitors (TKIs), such as Sunitinib and other TKIs, have been demonstrated to lower blood glucose levels in

T2D patients [89]. Sunitinib was administered to patients with KIT overexpression, and responses were evaluated using the Response Evaluation Criteria in Solid Tumors (RECIST) and those with KIT mutations are more responsive to Sunitinib. As well, sunitinib may be superior to other KIT inhibitors due to its ability to inhibit VEGFR1, VEGFR2, VEGFR3, and platelet-derived growth factor receptor (PDGFR) [90]. So, inhibiting mutant-like KIT in metastatic melanoma with nilotinib and dasatinib is becoming increasingly popular. Apart from binding and inhibiting ABL's kinase domain, nilotinib affects KIT and PDGF receptor kinases more effectively than imatinib [91]. In an accelerated-phase, CML patient with T2D, dasatinib reduced hyperglycemia rapidly [92]. The resulted drug is found from a shared analysis of T2D and smoking. Although the drug is only validated for T2D as per previous literature, it may also mediate the negative consequence that occurred from smoking. Therefore, our predicted drug may be well performed for T2D with a smoking history.

We have tried to validate all our findings with previous literature. As it was not possible, therefore, there is still scope for further in vivo and in vitro level research. Based on the findings, the doctor/physician may recommend that active smokers abstain from the bad habit to minimise the likelihood of T2D. Furthermore, smoking-controlled complications in T2D patients are a major concern for T2D treatment. Thus, avoiding smoking is highly recommended for T2D patients. Moreover, the pharmaceutical sectors may develop drugs depending on the resulted chemical compound for the treatment of T2D, which may also reduce the evolved smoking effect. For the limitations of this study, it should be noted that our selected datasets have fewer samples. Age, sex, ethnicity and other relevant attributes were not considered for this study. Hence, further validation is required to thoroughly analyse the biological relevance related to the stated problem in this study.

5. Conclusions

To obtain better knowledge about the risk of smoking, we have conducted a bioinformatics analysis in this study. We have found that smokers have a greater risk of developing Type 2 diabetes, and T2D patients could quickly evolve many complications due to smoking. Therefore, we have conducted various statistical analyses on both transcriptomic and GWAS profiles of active smokers and T2D patients. Our study resulted in 55 shared biomarker DEGs between smokers and T2D from the transcriptomic analysis. Whereas seven genes (SPP1, PSMB8, PRRC2A, PRIM1, KIT, NTRK1 and HLA-C) are validated using previous studies. They are also repeated in multiple analyses such as GWAS analysis, protein–protein interaction, protein–drug interactions, and transcriptional and post-transcriptional Regulators. Signalling pathways such as PI3K/AKT signalling pathway, AMPK signalling pathway, Endosomal/Vacuolar pathways and Allograft Rejection are also verified using published literature. However, still, there are many novel findings that make way for further analysis by the researchers. Finally, our study suggests that people should quit smoking as soon as possible in order to reduce the risk of being diagnosed with type 2 diabetes. Moreover, type 2 diabetes patients must quit smoking to avoid other complications/diseases and take medication accordingly.

Author Contributions: A.S.M.R.R. and M.A.A. have contribute equally and claim combined first author. Conceptualization, A.S.M.R.R., M.A.A. and M.K.I.; methodology, M.K.I.; software, M.A.A. and M.K.I.; validation, M.A.R., A.S.M.R.R., F.H. and K.R.A.; formal analysis, A.S.M.R.R., M.A.A., M.K.I., F.H. and K.R.A.; investigation, M.A.A., A.S.M.R.R. and M.K.I.; resources, A.S.M.R.R., M.A.A. and M.K.I.; data curation, M.A.A., A.S.M.R.R. and M.K.I.; writing—original draft preparation, A.S.M.R.R., M.A.A. and M.K.I.; writing—review and editing, M.A.A., M.K.I., A.S.M.R.R., F.H. and K.R.A.; visualization, M.A.A.; supervision, M.A.R., M.Z.I., and B.K.; project administration, M.A.R., M.Z.I. and B.K.; funding acquisition, M.A.R. and B.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Korea Institute of Oriental Medicine (grant number KSN2021240), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1I1A2066868), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A5A2019413), a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health&Welfare, Republic of Korea (grant number : HF20C0116), and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health&Welfare, Republic of Korea (grant number: HF20C0038).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Links of our dataset are included in this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Olokoba, A.B.; Obateru, O.A.; Olokoba, L.B. Type 2 diabetes mellitus: A review of current trends. *Oman Med. J.* **2012**, *27*, 269. [CrossRef] [PubMed]
2. Beta Cells. 2021. Available online: <https://www.medicalnewstoday.com/articles/beta-cells-in-type-2-diabetes> (accessed on 14 April 2022).
3. Martín-Timón, I.; Sevillano-Collantes, C.; Segura-Galindo, A.; del Cañizo-Gómez, F.J. Type 2 diabetes and cardiovascular disease: Have all risk factors the same strength? *World J. Diabetes* **2014**, *5*, 444. [CrossRef] [PubMed]
4. Cho, N.; Shaw, J.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J.; Ohlrogge, A.; Malanda, B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* **2018**, *138*, 271–281. [CrossRef]
5. Diabetic-WHO. 2021. Available online: <https://www.who.int/news-room/fact-sheets/detail/diabetes> (accessed on 14 April 2022).
6. Diabetes Smoking. 2022. Available online: <https://www.cdc.gov/tobacco/campaign/tips/diseases/diabetes.html> (accessed on 14 April 2022).
7. Xu, H.; Wang, Q.; Sun, Q.; Qin, Y.; Han, A.; Cao, Y.; Yang, Q.; Yang, P.; Lu, J.; Liu, Q.; et al. In type 2 diabetes induced by cigarette smoking, activation of p38 MAPK is involved in pancreatic β -cell apoptosis. *Environ. Sci. Pollut. Res.* **2018**, *25*, 9817–9827. [CrossRef]
8. Consortium, I.; Spijkerman, A.M.; van der A, D.L.; Nilsson, P.M.; Ardanaz, E.; Gavrila, D.; Agudo, A.; Arriola, L.; Balkau, B.; Beulens, J.W.; et al. Smoking and long-term risk of type 2 diabetes: The EPIC-InterAct study in European populations. *Diabetes Care* **2014**, *37*, 3164–3171. [CrossRef] [PubMed]
9. Yuan, S.; Xue, H.L.; Yu, H.J.; Huang, Y.; Tang, B.W.; Yang, X.H.; Li, Q.X.; He, Q.Q. Cigarette smoking as a risk factor for type 2 diabetes in women compared with men: A systematic review and meta-analysis of prospective cohort studies. *J. Public Health* **2019**, *41*, e169–e176. [CrossRef]
10. Östgren, C.J.; Lindblad, U.; Ranstam, J.; Melander, A.; Råstam, L. Associations between smoking and β -cell function in a non-hypertensive and non-diabetic population Skaraborg Hypertension and Diabetes Project. *Diabet. Med.* **2000**, *17*, 445–450. [CrossRef] [PubMed]
11. Liu, T.; Wang, H.; Qiu, Q.; Tan, L.L.; Chen, W.; Yu, X.Q.; Sun, X.L.; Chen, W.Q. Mediation of abdominal obesity on the association between cigarette smoking and β -cell function. *Zhonghua Liu Xing Bing Xue Za Zhi* **2010**, *31*, 988–991. [PubMed]
12. Artese, A.; Stamford, B.A.; Moffatt, R.J. Cigarette smoking: An accessory to the development of insulin resistance. *Am. J. Lifestyle Med.* **2019**, *13*, 602–605. [CrossRef]
13. Mouhamed, D.H.; Ezzaher, A.; Neffati, F.; Douki, W.; Gaha, L.; Najjar, M. Effect of cigarette smoking on insulin resistance risk. In *Proceedings of the Annales de Cardiologie et d'Angéiologie*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 65, pp. 21–25.
14. Ren, W.; Gu, Y.; Zhu, L.; Wang, L.; Chang, Y.; Yan, M.; Han, B.; He, J. The effect of cigarette smoking on vitamin D level and depression in male patients with acute ischemic stroke. *Compr. Psychiatry* **2016**, *65*, 9–14. [CrossRef]
15. Piazzolla, G.; Castrovilli, A.; Liotino, V.; Vulpi, M.R.; Fanelli, M.; Mazzocca, A.; Candigliota, M.; Berardi, E.; Resta, O.; Sabbà, C.; et al. Metabolic syndrome and Chronic Obstructive Pulmonary Disease (COPD): The interplay among smoking, insulin resistance and vitamin D. *PLoS ONE* **2017**, *12*, e0186708. [CrossRef] [PubMed]
16. Lips, P.; Eekhoff, M.; van Schoor, N.; Oosterwerff, M.; de Jongh, R.; Krul-Poel, Y.; Simsek, S. Vitamin D and type 2 diabetes. *J. Steroid Biochem. Mol. Biol.* **2017**, *173*, 280–285. [CrossRef] [PubMed]
17. Haase, C.L.; Tybjaerg-Hansen, A.; Nordestgaard, B.G.; Frikke-Schmidt, R. HDL cholesterol and risk of type 2 diabetes: A Mendelian randomization study. *Diabetes* **2015**, *64*, 3328–3333. [CrossRef] [PubMed]
18. Abbasi, A.; Corpeleijn, E.; Gansevoort, R.T.; Gans, R.O.; Hillege, H.L.; Stolk, R.P.; Navis, G.; Bakker, S.J.; Dullaart, R.P. Role of HDL cholesterol and estimates of HDL particle composition in future development of type 2 diabetes in the general population: The PREVEND study. *J. Clin. Endocrinol. Metab.* **2013**, *98*, E1352–E1359. [CrossRef]
19. Rader, D.J.; Hobbs, H.H. Disorders of lipoprotein metabolism. *Harrisons Princ. Intern. Med.* **2005**, *16*, 2286.

20. He, B.M.; Zhao, S.P.; Peng, Z.Y. Effects of cigarette smoking on HDL quantity and function: Implications for atherosclerosis. *J. Cell. Biochem.* **2013**, *114*, 2431–2436. [[CrossRef](#)]
21. Zhao, J.; Zhang, Y.; Wei, F.; Song, J.; Cao, Z.; Chen, C.; Zhang, K.; Feng, S.; Wang, Y.; Li, W.D. Triglyceride is an independent predictor of type 2 diabetes among middle-aged and older adults: A prospective study with 8-year follow-ups in two cohorts. *J. Transl. Med.* **2019**, *17*, 1–7. [[CrossRef](#)]
22. Bading-Taïka, B.; Souza, A.; Bourobou Bourobou, H.P.; Lione, L.A. Hypoglycaemic and anti-hyperglycaemic activity of Tabernanthe iboga aqueous extract in fructose-fed streptozotocin type 2 diabetic rats. *Adv. Tradit. Med.* **2021**, *21*, 281–295. [[CrossRef](#)]
23. Koda, M.; Kitamura, I.; Okura, T.; Otsuka, R.; Ando, F.; Shimokata, H. The associations between smoking habits and serum triglyceride or hemoglobin A1c levels differ according to visceral fat accumulation. *J. Epidemiol.* **2016**, *26*, JE20150064. [[CrossRef](#)]
24. Endocrinology. 2022. Available online: <https://rb.gy/eptoy1> (accessed on 13 June 2022).
25. Anjum, A.; Jaggi, S.; Varghese, E.; Lall, S.; Bhowmik, A.; Rai, A. Identification of differentially expressed genes in rna-seq data of arabidopsis thaliana: A compound distribution approach. *J. Comput. Biol.* **2016**, *23*, 239–247. [[CrossRef](#)]
26. Mahi, N.A.; Najafabadi, M.F.; Pilarczyk, M.; Kouril, M.; Medvedovic, M. GREIN: An interactive web platform for re-analyzing GEO RNA-seq data. *Sci. Rep.* **2019**, *9*, 7580. [[CrossRef](#)] [[PubMed](#)]
27. Kovács, S.A.; Györfy, B. Transcriptomic datasets of cancer patients treated with immune-checkpoint inhibitors: A systematic review. *J. Transl. Med.* **2022**, *20*, 249. [[CrossRef](#)] [[PubMed](#)]
28. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
29. Vandereyken, K.; Van Leene, J.; De Coninck, B.; Cammue, B. Hub protein controversy: Taking a closer look at plant stress response hubs. *Front. Plant Sci.* **2018**, *9*, 694. [[CrossRef](#)]
30. Ernst, J.; Beg, Q.K.; Kay, K.A.; Balázs, G.; Oltvai, Z.N.; Bar-Joseph, Z. A semi-supervised method for predicting transcription factor–gene interactions in Escherichia coli. *PLoS Comput. Biol.* **2008**, *4*, e1000044. [[CrossRef](#)]
31. Haussler, J.; Zavolan, M. Identification and consequences of miRNA–target interactions—Beyond repression of gene expression. *Nat. Rev. Genet.* **2014**, *15*, 599–612. [[CrossRef](#)]
32. Miyagi, M.; Tanaka, K.; Watanabe, S.; Kondo, J.; Kishimoto, T. Identifying Protein–Drug Interactions in Cell Lysates Using Histidine Hydrogen Deuterium Exchange. *Anal. Chem.* **2021**, *93*, 14985–14995. [[CrossRef](#)]
33. Jia, P.; Wang, L.; Meltzer, H.Y.; Zhao, Z. Common variants conferring risk of schizophrenia: A pathway analysis of GWAS data. *Schizophr. Res.* **2010**, *122*, 38–42. [[CrossRef](#)]
34. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)]
35. Islam, M.K.; Rahman, M.H.; Islam, M.R.; Islam, M.Z.; Mamun, M.M.I.; Azad, A.; Moni, M.A. Network based systems biology approach to identify disease and comorbidity associations of Systemic Sclerosis with cancers. *Heliyon* **2022**, *8*, e08892. [[CrossRef](#)]
36. Spallotta, F.; Cencioni, C.; Atlante, S.; Garella, D.; Cocco, M.; Mori, M.; Mastrocola, R.; Kuenne, C.; Guenther, S.; Nanni, S.; et al. Stable oxidative cytosine modifications accumulate in cardiac mesenchymal cells from type2 diabetes patients: Rescue by α -ketoglutarate and TET-TDG functional reactivation. *Circ. Res.* **2018**, *122*, 31–46. [[CrossRef](#)] [[PubMed](#)]
37. Ryan, D.M.; Vincent, T.L.; Salit, J.; Walters, M.S.; Agosto-Perez, F.; Shaykhiev, R.; Strulovici-Barel, Y.; Downey, R.J.; Buro-Auriemma, L.J.; Staudt, M.R.; et al. Smoking dysregulates the human airway basal cell transcriptome at COPD risk locus 19q13. 2. *PLoS ONE* **2014**, *9*, e88051. [[CrossRef](#)] [[PubMed](#)]
38. Rahman, M.H.; Peng, S.; Hu, X.; Chen, C.; Rahman, M.R.; Uddin, S.; Quinn, J.M.; Moni, M.A. A network-based bioinformatics approach to identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1035. [[CrossRef](#)] [[PubMed](#)]
39. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [[CrossRef](#)]
40. Gene Ontology Consortium. Creating the gene ontology resource: Design and implementation. *Genome Res.* **2001**, *11*, 1425–1433. [[CrossRef](#)]
41. García-Campos, M.A.; Espinal-Enríquez, J.; Hernández-Lemus, E. Pathway analysis: State of the art. *Front. Physiol.* **2015**, *6*, 383. [[CrossRef](#)]
42. Ramos, E.M.; Hoffman, D.; Junkins, H.A.; Maglott, D.; Phan, L.; Sherry, S.T.; Feolo, M.; Hindorff, L.A. Phenotype–Genotype Integrator (PheGenI): Synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur. J. Hum. Genet.* **2014**, *22*, 144–147. [[CrossRef](#)]
43. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2019**, *47*, D1005–D1012. [[CrossRef](#)]

44. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **2016**, *45*, gkw937. [[CrossRef](#)]
45. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
46. Chen, S.H.; Chin, C.H.; Wu, H.H.; Ho, C.W.; Ko, M.T.; Lin, C.Y. cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology. In Proceedings of the 20th International Conference on Genome Informatics, Yokohama, Japan, 14–16 December 2009.
47. Xia, J.; Gill, E.E.; Hancock, R.E. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* **2015**, *10*, 823–844. [[CrossRef](#)] [[PubMed](#)]
48. Cheng, C.; Alexander, R.; Min, R.; Leng, J.; Yip, K.Y.; Rozowsky, J.; Yan, K.K.; Dong, X.; Djebali, S.; Ruan, Y.; et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* **2012**, *22*, 1658–1667. [[CrossRef](#)] [[PubMed](#)]
49. Jonker, N.; Kool, J.; Irth, H.; Niessen, W. Recent developments in protein–ligand affinity mass spectrometry. *Anal. Bioanal. Chem.* **2011**, *399*, 2669–2681. [[CrossRef](#)]
50. de Azevedo, J.; Walter, F.; Caceres, R.A.; Pauli, I.; Timmers, L.F.S.; Barcellos, G.B.; Rocha, K.B.; Soares, M.B. Protein-drug interaction studies for development of drugs against Plasmodium falciparum. *Curr. Drug Targets* **2009**, *10*, 271–278. [[CrossRef](#)] [[PubMed](#)]
51. Lachmann, A.; Xu, H.; Krishnan, J.; Berger, S.I.; Mazloom, A.R.; Ma’ayan, A. ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **2010**, *26*, 2438–2444. [[CrossRef](#)] [[PubMed](#)]
52. Khan, A.; Fornes, O.; Stigliani, A.; Gheorghe, M.; Castro-Mondragon, J.A.; Van Der Lee, R.; Bessy, A.; Cheneby, J.; Kulkarni, S.R.; Tan, G.; et al. JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **2018**, *46*, D260–D266. [[CrossRef](#)]
53. Sethupathy, P.; Corda, B.; Hatzigeorgiou, A.G. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna* **2006**, *12*, 192–197. [[CrossRef](#)]
54. Hsu, S.D.; Lin, F.M.; Wu, W.Y.; Liang, C.; Huang, W.C.; Chan, W.L.; Tsai, W.T.; Chen, G.Z.; Lee, C.J.; Chiu, C.M.; et al. miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **2011**, *39*, D163–D169. [[CrossRef](#)]
55. Wang, C.; Hu, G.; Wang, K.; Brylinski, M.; Xie, L.; Kurgan, L. PDID: Database of molecular-level putative protein–drug interactions in the structural human proteome. *Bioinformatics* **2016**, *32*, 579–586. [[CrossRef](#)]
56. Miao, T.W.; Xiao, W.; Du, L.y.; Mao, B.; Huang, W.; Chen, X.m.; Li, C.; Wang, Y.; Fu, J.j. High expression of SPP1 in patients with chronic obstructive pulmonary disease (COPD) is correlated with increased risk of lung cancer. *FEBS Open Bio* **2021**, *11*, 1237–1249. [[CrossRef](#)]
57. Stoynev, N.; Dimova, I.; Rukova, B.; Hadjidekova, S.; Nikolova, D.; Toncheva, D.; Tankova, T. Gene expression in peripheral blood of patients with hypertension and patients with type 2 diabetes. *J. Cardiovasc. Med.* **2014**, *15*, 702–709. [[CrossRef](#)] [[PubMed](#)]
58. Tichelaar, J.W.; Borchers, M.T.; Wesselkamper, S.C.; Curull, V.; Ramirez-Sarmiento, A.; Sanchez-Font, A.; Garcia-Aymerich, J.; Coronell, C.; Lloreta, J.; Agusti, A.G.; et al. Sustained CTL activation by murine pulmonary epithelial cells promotes the development of COPD-like disease. *J. Clin. Investig.* **2009**, *119*, 636–649.
59. Prince, C.; Hammerton, G.; Taylor, A.E.; Anderson, E.L.; Timpson, N.J.; Davey Smith, G.; Munafò, M.R.; Relton, C.L.; Richmond, R.C. Investigating the impact of cigarette smoking behaviours on DNA methylation patterns in adolescence. *Hum. Mol. Genet.* **2019**, *28*, 155–165. [[CrossRef](#)]
60. Rai, A.; Pawar, A.K.; Jalan, S. Prognostic interaction patterns in diabetes mellitus II: A random-matrix-theory relation. *Phys. Rev. E* **2015**, *92*, 022806. [[CrossRef](#)]
61. Li, Y.; Yuan, X.; Shi, Z.; Wang, H.; Ren, D.; Zhang, Y.; Fan, Y.; Liu, Y.; Cui, Z. LncRNA XIST serves as a diagnostic biomarker in gestational diabetes mellitus and its regulatory effect on trophoblast cell via miR-497-5p/FOXO1 axis. *Cardiovasc. Diagn. Ther.* **2021**, *11*, 716. [[CrossRef](#)]
62. Kaneko, K.; Katagiri, H. Dual-specificity phosphatase 8: A gatekeeper in hypothalamic control of glucose metabolism in males. *J. Diabetes Investig.* **2021**, *12*, 1138. [[CrossRef](#)] [[PubMed](#)]
63. Abo-Elfetoh, N.M.; Mohamed, E.S.; Tag, L.M.; Gamal, R.M.; Gandour, A.M.; Razek, A.E.; Mohamed, R.; El-Baz, M.A.; Ez Eldeen, M.E. The relationship between auditory brainstem response, nerve conduction studies, and metabolic risk factors in type II diabetes mellitus. *Egypt. Rheumatol. Rehabil.* **2016**, *43*, 163–171. [[CrossRef](#)]
64. Liu, M.; Xu, K.; Saaoud, F.; Shao, Y.; Zhang, R.; Lu, Y.; Sun, Y.; Drummer, C.; Li, L.; Wu, S.; et al. 29 m6A-RNA Methylation (Epitranscriptomic) Regulators Are Regulated in 41 Diseases including Atherosclerosis and Tumors Potentially via ROS Regulation–102 Transcriptomic Dataset Analyses. *J. Immunol. Res.* **2022**, *2022*, 1433323. [[CrossRef](#)]
65. Laqqan, M.M.; Yassin, M.M. Influence of tobacco cigarette heavy smoking on DNA methylation patterns and transcription levels of MAPK8IP3, GAA, ANXA2, PRRC2A, and PDE11A genes in human spermatozoa. *Middle East Fertil. Soc. J.* **2021**, *26*, 41. [[CrossRef](#)]
66. Ma, L.; Hanson, R.L.; Que, L.N.; Cali, A.M.; Fu, M.; Mack, J.L.; Infante, A.M.; Kobes, S.; The International Type 2 Diabetes Consortium; Bogardus, C.; et al. Variants in ARHGEF11, a candidate gene for the linkage to type 2 diabetes on chromosome 1q, are nominally associated with insulin resistance and type 2 diabetes in Pima Indians. *Diabetes* **2007**, *56*, 1454–1459. [[CrossRef](#)]

67. Wang, J.; Kataoka, H.; Suzuki, M.; Sato, N.; Nakamura, R.; Tao, H.; Maruyama, K.; Isogaki, J.; Kanaoka, S.; Ihara, M.; et al. Downregulation of EphA7 by hypermethylation in colorectal cancer. *Oncogene* **2005**, *24*, 5637–5647. [[CrossRef](#)] [[PubMed](#)]
68. Stavarachi, M.; Panduru, N.; Serafinceanu, C.; Moța, E.; Moța, M.; Cimponeriu, D.; Ion, D.A. Investigation of P213S SELL gene polymorphism in type 2 diabetes mellitus and related end stage renal disease. A case-control study. *Rom J. Morphol. Embryol.* **2011**, *52*, 995–998. [[PubMed](#)]
69. Zhu, M.; Wu, M.; Bian, S.; Song, Q.; Xiao, M.; Huang, H.; You, L.; Zhang, J.; Zhang, J.; Cheng, C.; et al. DNA primase subunit 1 deteriorated progression of hepatocellular carcinoma by activating AKT/mTOR signaling and UBE2C-mediated P53 ubiquitination. *Cell Biosci.* **2021**, *11*, 42. [[CrossRef](#)]
70. Lee, W.H.; Chen, L.C.; Lee, C.J.; Huang, C.C.; Ho, Y.S.; Yang, P.S.; Ho, C.T.; Chang, H.L.; Lin, I.H.; Chang, H.W.; et al. DNA primase polypeptide 1 (PRIM1) involves in estrogen-induced breast cancer formation through activation of the G2/M cell cycle checkpoint. *Int. J. Cancer* **2019**, *144*, 615–630. [[CrossRef](#)]
71. Hammouz, R.Y.; Kostanek, J.K.; Dudzisz, A.; Witas, P.; Orzechowska, M.; Bednarek, A.K. Differential expression of lung adenocarcinoma transcriptome with signature of tobacco exposure. *J. Appl. Genet.* **2020**, *61*, 421–437. [[CrossRef](#)]
72. Ambatipudi, S.; Cuenin, C.; Hernandez-Vargas, H.; Ghantous, A.; Le Calvez-Kelm, F.; Kaaks, R.; Barrdahl, M.; Boeing, H.; Aleksandrova, K.; Trichopoulou, A.; et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* **2016**, *8*, 599–618. [[CrossRef](#)] [[PubMed](#)]
73. Wu, C.; Xu, G.; Tsai, S.Y.A.; Freed, W.J.; Lee, C.T. Transcriptional profiles of type 2 diabetes in human skeletal muscle reveal insulin resistance, metabolic defects, apoptosis, and molecular signatures of immune activation in response to infections. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 282–288. [[CrossRef](#)]
74. Lin, F.; Yang, D.; Huang, Y.; Zhao, Y.; Ye, J.; Xiao, M. The potential of neoagaro-oligosaccharides as a treatment of type II diabetes in mice. *Mar. Drugs* **2019**, *17*, 541. [[CrossRef](#)]
75. Sandhu, H.; Xu, C.B.; Edvinsson, L. Upregulation of contractile endothelin type B receptors by lipid-soluble cigarette smoking particles in rat cerebral arteries via activation of MAPK. *Toxicol. Appl. Pharmacol.* **2010**, *249*, 25–32. [[CrossRef](#)]
76. Thomas, M.C.; Mathew, T.; Russ, G.; Rao, M.M.; Moran, J. Early peri-operative glycaemic control and allograft rejection in patients with diabetes mellitus: A pilot study. *Transplantation* **2001**, *72*, 1321–1324. [[CrossRef](#)]
77. Ohiomoba, R.; Youmans, Q.R.; Ezema, A.; Anderson, A.S.; Jackson, K.; Mandieka, E.; Pham, D.T.; Rich, J.D.; Yancy, C.W.; Okwuosa, I.S. History of Cigarette Smoking and Heart Transplant Outcomes. *J. Card. Fail.* **2019**, *25*, S182. [[CrossRef](#)]
78. Khanna, A.K.; Xu, J.; Uber, P.A.; Burke, A.P.; Baquet, C.; Mehra, M.R. Tobacco smoke exposure in either the donor or recipient before transplantation accelerates cardiac allograft rejection, vascular inflammation, and graft loss. *Circulation* **2009**, *120*, 1814–1821. [[CrossRef](#)] [[PubMed](#)]
79. Huang, X.; Liu, G.; Guo, J.; Su, Z. The PI3K/AKT pathway in obesity and type 2 diabetes. *Int. J. Biol. Sci.* **2018**, *14*, 1483. [[CrossRef](#)] [[PubMed](#)]
80. Sun, X.; Chen, L.; He, Z. PI3K/Akt-Nrf2 and anti-inflammation effect of macrolides in chronic obstructive pulmonary disease. *Curr. Drug Metab.* **2019**, *20*, 301–304. [[CrossRef](#)] [[PubMed](#)]
81. Wei, K.W. *Identification of Shared Molecular Pathways and Networks between Alzheimer's Disease and Type 2 Diabetes*; University of California: Los Angeles, CA, USA, 2016.
82. Žemeckienė, Ž.; Vitkauskienė, A.; Šjakste, T.; Šitkauskienė, B.; Sakalauskas, R. Proteasomes and proteasomal gene polymorphism in association with inflammation and various diseases. *Medicina* **2013**, *49*, 33. [[CrossRef](#)]
83. Fjukstad, K.; Athanasiu, L.; Dieset, I.; Steen, N.; Djurovic, S.; Spigset, O.; Andreassen, O. Metabolic Abnormalities Related to Treatment with Selective Serotonin Reuptake Inhibitors in Patients with Schizophrenia or Bipolar Disorder: A Genome Wide Association Study. *Eur. Neuropsychopharmacol.* **2019**, *29*, S998–S999. [[CrossRef](#)]
84. Zhang, C.; Shen, L.; Zhu, Y.; Xu, R.; Deng, Z.; Liu, X.; Ding, Y.; Wang, C.; Shi, Y.; Bei, L.; et al. KDM6A promotes imatinib resistance through YY1-mediated transcriptional upregulation of TRKA independently of its demethylase activity in chronic myelogenous leukemia. *Theranostics* **2021**, *11*, 2691. [[CrossRef](#)]
85. Dingli, D.; Wolf, R.C.; Vella, A. Imatinib and type 2 diabetes. *Endocr. Pract.* **2007**, *13*, 126–130. [[CrossRef](#)]
86. Imarisio, I.; Paglino, C.; Ganini, C.; Magnani, L.; Caccialanza, R.; Porta, C. The effect of sorafenib treatment on the diabetic status of patients with renal cell or hepatocellular carcinoma. *Future Oncol.* **2012**, *8*, 1051–1057. [[CrossRef](#)]
87. Guo, T.; Agaram, N.P.; Wong, G.C.; Hom, G.; D'Adamo, D.; Maki, R.G.; Schwartz, G.K.; Veach, D.; Clarkson, B.D.; Singer, S.; et al. Sorafenib inhibits the imatinib-resistant KIT T670I gatekeeper mutation in gastrointestinal stromal tumor. *Clin. Cancer Res.* **2007**, *13*, 4874–4881. [[CrossRef](#)]
88. Tyrrell, H.; Pwint, T. Sunitinib and improved diabetes control. *BMJ Case Rep.* **2014**, *2014*, bcr2014207521. [[CrossRef](#)]
89. Huda, M.S.; Amiel, S.A.; Ross, P.; Aylwin, S.J. Tyrosine kinase inhibitor sunitinib allows insulin independence in long-standing type 1 diabetes. *Diabetes Care* **2014**, *37*, e87–e88. [[CrossRef](#)] [[PubMed](#)]
90. Minor, D.R.; Kashani-Sabet, M.; Garrido, M.; O'Day, S.J.; Hamid, O.; Bastian, B.C. Sunitinib Therapy for Melanoma Patients with KIT Mutations Sunitinib Therapy for Melanoma with KIT Mutations. *Clin. Cancer Res.* **2012**, *18*, 1457–1463. [[CrossRef](#)] [[PubMed](#)]

-
91. Tran, A.; Tawbi, H.A. A potential role for nilotinib in KIT-mutated melanoma. *Expert Opin. Investig. Drugs* **2012**, *21*, 861–869. [[CrossRef](#)] [[PubMed](#)]
 92. Ono, K.; Suzushima, H.; Watanabe, Y.; Kikukawa, Y.; Shimomura, T.; Furukawa, N.; Kawaguchi, T.; Araki, E. Rapid amelioration of hyperglycemia facilitated by dasatinib in a chronic myeloid leukemia patient with type 2 diabetes mellitus. *Intern. Med.* **2012**, *51*, 2763–2766. [[CrossRef](#)]