



## Probing biological network in concurrent carcinomas and Type-2 diabetes for potential biomarker screening: An advanced computational paradigm

Abdullah Al Marzan<sup>1</sup>, Shatila Shahi<sup>1</sup>, Md Sakil Arman<sup>1</sup>, Md Zafrul Hasan,  
Ajit Ghosh\*

Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet, 3114, Bangladesh



### ARTICLE INFO

#### Keywords:

Type-2 diabetes mellitus  
Bladder-cancers  
Breast-cancer  
Prostate-cancer  
Liver-cancer  
Colon-cancer  
Co-morbidity  
Biomarker  
Biological-network

### ABSTRACT

Type-2 diabetes mellitus (T2DM), the predominant form of diabetes in adults, is a co-morbid condition that exacerbates the severity of many other diseases, including cardiovascular disease, obesity, dyslipidemia, hypertension, and cancer. Among these, cancer is particularly concerning due to elevated mortality rates and a distinct lack of cost-effective therapeutic interventions. Identifying novel biomarkers for improved early cancer detection is imperative. Therefore, an integrated bioinformatics analysis was conducted to elucidate the co-morbid relationship between T2DM and five different types of cancer, namely bladder (BLCA), breast (BRCA), colon (CRC), liver (HCC), and prostate cancer (PRAD) and identification of novel biomarkers for early cancer detection in individuals with T2DM. A significant comorbid relationship was observed among T2DM, BLCA, and BRCA through gene expression and pathway enrichment analysis, while a moderate association was observed for between T2DM, and PRAD. Notably, we identified 18 significant hub proteins in the context of cancer and T2DM, along with 16 transcription factors and 5 miRNAs. Among these, the hub proteins ESR1, PIK3CA, GNAI1, ERBB2, NR3C1, SNCA, TGFBR2, as well as the micro RNAs hsa-mir-335-5p, hsa-mir-16-5p, and hsa-mir-93-5p hold promise for understanding the comorbidities of T2DM and cancers; and could serve as valuable disease biomarkers for clinical diagnosis and prognosis. This study, centred on bioinformatics analysis for biomarker identification in comorbidities, paves the way for future research encompassing wet lab experimentation and translational studies. These endeavours are poised to validate and facilitate the integration of these findings into the realm of personalized medicine.

### 1. Introduction

Type-2 Diabetes Mellitus (T2DM) is the most frequent form of Diabetes Mellitus in the modern world, affecting approximately 463 million people globally in 2019, with a projected increase to 700.2 million by 2045.<sup>1</sup> The pathogenesis of T2DM involves insulin deficiency, insulin resistance, and improper insulin response by cells,<sup>2–4</sup> leading to a cellular microenvironment that promotes cancer growth, as reviewed by Tudzarova et al.<sup>5</sup> Cancer is attributed to the leading cause of global mortality, causing 10 million deaths in

\* Corresponding author.

E-mail address: [aghosh-bmb@sust.edu](mailto:aghosh-bmb@sust.edu) (A. Ghosh).

<sup>1</sup> Author Contributed Equally.

2020, with diabetic individuals having higher mortality rates.<sup>5,6</sup> Cancer mortality is expected to rise to 16 million by 2040.<sup>7,8</sup> The hypothesis linking cancer and diabetes has existed for over a century,<sup>9</sup> but it has gained widespread acceptance in recent years. Consequently, researchers are conducting population-based studies<sup>10,11</sup> to investigate the relationship between cancer and diabetes in different aspects, identifying associated risk factors<sup>12</sup> and exploring its impact on cancer diagnosis and treatment.<sup>13</sup>

Although the precise underlying mechanism between T2DM and cancer is not yet fully understood, several biological phenomena linking both conditions have been identified.<sup>14</sup> These include genetic predisposition,<sup>15</sup> as well as shared risk factors such as obesity,<sup>16</sup> hyperglycemia,<sup>17</sup> hyperinsulinemia,<sup>18</sup> and oxidative stress.<sup>19</sup> Obesity,<sup>10</sup> a common risk factor for multiple morbidities, is associated with stimulated cellular proliferation, hormonal imbalance, and inflammation, bridging T2DM and cancers.<sup>20</sup> Metabolic dysregulation in obesity is caused by adipose tissue inflammation in which noncanonical Wnt signalling, specifically wnt5a, plays a significant role.<sup>21</sup> Further, the dysregulated Wnt signalling promotes cancer cell proliferation<sup>18</sup> by altering glycolysis metabolism<sup>22</sup> while simultaneously interfering with adipogenesis.<sup>23</sup> These dysregulated phenomena ultimately lead to diabetes and cancer, interconnected in a molecular orchestra.<sup>24</sup>

In diabetic individuals, insulin resistance is countered by adaptive beta cell expansion to mitigate hyperglycemia, resulting in hyperinsulinemia.<sup>24</sup> Hyperinsulinemia can cause an imbalance in cell proliferation by overstimulating insulin and insulin-like growth factor (IGF) receptors or inhibiting insulin-like growth factor receptor binding proteins (IGFBPs), resulting in cancer in the lungs and other organs.<sup>25,26</sup> Enhanced levels of bioactive estrogens could be contributing to these interactions.<sup>27</sup> On the other hand, hyperglycemia has been attributed to the progression of various malignancies, including CRC, HCC, lung, and pancreatic cancer.<sup>17,28</sup> The cancer cells tend to have an increased demand for glucose for glycolysis due to the Warburg effect,<sup>29</sup> which contributes to a hyperglycemic state associated with diabetes, promoting immunosuppression and carcinogenesis.<sup>30</sup> In another mechanism, hyperglycemia-induced superoxide production has been found to influence gene expression and signal transduction associated with tumorigenesis.<sup>31–33</sup> Increased ROS production due to defective insulin secretion in T2DM activates pro-inflammatory pathways and adipokine production, which promotes tumorigenesis and metastasis.<sup>34</sup> Furthermore, ROS can interact with NF-κB pathways, altering apoptosis and cell proliferation,<sup>35</sup> and this interaction has been reported to be hyperactivated in CRC, BRCA, and pancreatic cancers.<sup>36</sup>

While various epidemiological and cohort studies have established a clinically important relationship between T2DM and various forms of malignancies, including Bladder Cancer (BLCA),<sup>37</sup> Breast Cancer (BRCA),<sup>38,39</sup> Colorectal cancer (CRC),<sup>40,41</sup> Hepatocellular carcinoma (HCC),<sup>42</sup> and Prostate Adenocarcinoma (PRAD),<sup>43,44</sup> the molecular puzzle between diabetic and individual cancer has not completely deciphered yet. Therefore, by utilizing advances in sequencing, bioinformatics, and available public databases, we can investigate a molecular interaction network that underlines the aetiology of the diseases.

Renzi et al. highlighted in their review the impact of multiple morbidities on the delayed diagnosis of cancer, primarily due to the complex and elusive relationship between these diseases, posing significant challenges in treatment.<sup>13</sup> Therefore, our study aimed to determine the biological connection between T2DM and prevalent cancers such as BLCA, BRCA, CRC, HCC, and PRAD through integrated bioinformatics analysis, identifying their comorbid mechanisms while finding significant biomarkers with diagnostic value. Therefore, we conducted a comprehensive analysis by mining publicly available data on selected cancer and T2DM from the NCBI GEO database.<sup>45</sup> We performed differential gene expression analysis, pathway enrichment analysis, and protein-protein interaction and miRNA-mRNA interaction analyses using several bioinformatics tools and databases such as Geo2R,<sup>45</sup> EnrichR,<sup>46</sup> NetworkAnalyst,<sup>47</sup> Cytoscape<sup>48</sup> and others. Our analysis identified commonly expressed genes that shared molecular pathways, 18 significant hub proteins, 16 transcription factors, and five microRNAs. These findings could be explored further to elucidate the intricate relationship between T2DM and cancer. Furthermore, we confirmed the significance of the identified biomarkers, particularly the hub proteins, by analyzing their levels in various tumour stages using ULCAN cancer databases.<sup>49,50</sup> We also analyzed cancer dependency<sup>51,52</sup> of these hub genes. The hub proteins identified by our research hold promise as potential biomarkers and therapeutic targets for cancer diagnosis and treatment in diabetic patients, whereas miRNA can be used to modify oncogene expression. However, our study utterly depends on a computational approach, and wet lab validation is required to understand our findings *in vivo*. Given the instrumental role of cancer biomarkers in guiding individualized cancer treatment,<sup>53</sup> our findings will contribute to the advancement of cancer diagnostics worldwide, especially in the context of personalized and precision medicine.

## 2. Material and methods

### 2.1. Data mining

Datasets for T2DM and different types of cancers were retrieved from NCBI Gene Expression Omnibus (GEO).<sup>54</sup> For each query, it was ensured that the datasets were structured to compare the diseased sample to the control group, that they were repeatable, and that only human data were included. Only microarray datasets with a sample size greater than 10 were evaluated for this investigation to ensure that each sample group has at least three replicates, thus supporting data reliability. The accession of selected datasets included GSE25724, GSE7476, GSE21422, GSE41328, GSE46408, and GSE38241 for T2DM, BLCA, BRCA, CRC, HCC, and PRAD, respectively.

T2DM (GSE25724) dataset contained 13 samples of pancreatic islets, including seven control samples from non-diabetic patients (age  $58 \pm 17$  years; gender, 4 males/3 females; body mass index,  $24.8 \pm 2.5$  kg/m<sup>2</sup>) and six T2DM patient (age  $71 \pm 9$  year; gender, 3 males/3 females; body mass index,  $26.0 \pm 2.2$  kg/m<sup>2</sup>) samples performed using Affymetrix Human Genome U133A Array.<sup>55</sup> The BLCA (GSE7476) dataset from Affymetrix Human Genome U133 Plus 2.0 Array consisted of 12 samples, including three samples of normal bladder tissue as controls, three samples of low-grade superficial tumours, three samples of high-grade superficial tumours with unclear clinical behaviour, and three samples of high-grade muscle-invasive tumours.<sup>56</sup> Healthy bladder tissue was collected from urothelial samples of 12 patients with no evidence of bladder malignancy. The collected cancerous samples are divided into tumour

stage TaGL(low grade, non-invasive), T1GH(high grade), T2-4(high grade, muscle-invasive, stage, T2, T3, T4). On the other hand, the BRCA (GSE21422) dataset, retrieved from the same platform, was generated with 5 healthy tissue samples as controls, 9 Ductal carcinomas in situ (DCIS), and 5 invasive ductal carcinomas.<sup>57</sup> The replicates were made from nineteen freshly frozen human breast tumour samples and five healthy control samples obtained from patients with breast reduction surgery. The average age of the female patients was 63 years. It was ensured that they were all free of distant metastasis.

The CRC dataset (GSE41328) included five normal colonic tissues and five colorectal adenocarcinomas that were evaluated with Affymetrix HG-U133-Plus-2.0 microarrays. Two laboratories independently generated microarray data on the same biological samples using the same array platform.<sup>58</sup> HCC (GSE46408) dataset derived from Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F includes mRNA expression pattern of 6 primary hepatocellular carcinomas and their associated nontumorous liver parenchyma.<sup>59</sup> The replicates were made from biological samples donated by 201 patients including 157 men and 44 women with a mean age of 55.1 years. Multiple anatomically diverse metastases from five patients and 21 normal prostate samples from organ donors were used to create the PRAD (GSE38241) dataset on the Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F platform.<sup>60</sup> Each sample was ensured to be devoid of any laboratory treatment (like depletion of any gene, treatment with any drug to see drug effect, etc.) for each disease. These datasets were derived through extensive screening to ensure that the differential expression focuses on giving a comparative view of healthy and diseased samples. This perspective would give us enough genes for further analysis, reducing statistical biases.

## 2.2. Quality assessment

Each dataset was examined for previous background correction and normalization methods from the corresponding sample papers to proceed further. GSE25724 and GSE7476 were normalized with RMA (robust multiarray average), while for GSE21422, GCRMA (GC robust multiarray average) was used to perform background correction and normalization. GSE46408 and GSE41328 followed the quantile normalization method. GSE38241 was normalized with loess and quantile method followed by background correction with normexp\_model by the Bioconductor limma (v 3.10.3). For GSE6408, we applied the quantile normalization method to ensure comparable gene expression levels across different samples. This is done using the “Force Normalization” option provided by GEO2R (an interactive web tool comparing two or more groups of samples in a GEO Series to identify differentially expressed genes). The expression data presented in the boxplots provided by GEO2R was observed (Fig. S1) to identify any potential outliers.

We also conducted principal component analysis (PCA) in R (version 4.3.0) with the FactoMineR (version 2.8) package. FactoMineR is an R package that provides a set of functions for multivariate exploratory data analysis (EDA) and dimensionality reduction techniques that are used for analyzing and visualizing high-dimensional data sets, such as gene expression data, metabolomics data, and survey data. On the other hand, PCA is used to reduce the complexity of high-dimensional gene expression data sets, identify outliers, detect potential clusters, and visualize relationships between samples. To identify distinct clusters and sample similarities, the K-mean clustering method followed by PCA analysis was conducted. The value of k was set to 2, 4, 3, 4, 2, 2 based on the sample group of GSE25724, GSE7476, GSE21422, GSE41328, GSE46408, GSE38241 accordingly. All the codes will be available upon request.

## 2.3. Additional datasets

To check the validity of our selected datasets, we also collected some additional datasets for each disease from ArrayExpress.<sup>61</sup> ArrayExpress is a database and resource hosted and maintained by the European Bioinformatics Institute (EBI), part of the European Molecular Biology Laboratory (EMBL). It stores and shares functional genomics experiments, including gene expression data derived from microarray and high-throughput sequencing technologies. We selected six datasets from this platform based on most studies similar to the datasets from GEO. Accession id for each dataset are, E-GEOD-24152,<sup>62</sup> E-GEOD-5764,<sup>63</sup> E-GEOD-50117,<sup>64</sup> E-GEOD-6222,<sup>65</sup> E-GEOD-20966,<sup>66</sup> E-GEOD-30994<sup>57</sup> for BLCA, BRCA, CRC, HCC, T2DM, and PRAD, respectively. Datasets for BLCA, BRCA, HCC and T2DM were performed with Affymetrix GeneChip, whereas the other two were performed with Agilent gene chips. For BRCA, 10 cancer samples were presented against 7 control samples from the bladder of individuals undergoing resection for non-urothelial carcinoma. For BRCA, samples were collected from the mammary glands of post-menopausal women. The samples were divided into regions normal ductal, tumor ductal, normal lobular, and tumor lobular, which contain 10, 5, 10, and 5 replicates, respectively. For CRC, 9 paired tumor-normal colorectal samples were analyzed and collected from both males and females (five from males, 4 from females). To construct HCC dataset, total RNA was extracted from human liver cancer at various stages (T1-1~4, T3-1~6). This dataset also contains 2 normal human livers and HuH7 cell line. For T2DM, the gene expression profile was constructed from samples of beta-cells obtained from 10 control and 10 T2DM patients. Healthy individuals (6 male and 4 female) were of age: 60 ± 5 years, BMI: 30.5 ± 6.5, while the patients (7 male and 3 female) were with age: 67 ± 7 and BMI: 30.9 ± 6.2 with a follow up of the disease for about 5.3 ± 2.3 years. PRAD dataset study was performed in normal prostate and cancerous prostate samples of age 59 ± 13.

Each dataset was normalized with Robust Multi-array Average (RMA) and outliers were reduced with the Z score method which calculates the Z-score for each data point in a dataset and removes any points that fall outside a certain threshold, typically set at around ±3 standard deviations from the mean. Therefore, Differential gene expression analysis was performed using limma (version 3.56.1) performing Benjamini & Hochberg's false discovery rate method. Finally, significant genes ( $p\text{-value} \leq 0.05$ ) were selected through the "dplyr" package (version 1.0.6) in R (version 4.0.4).

## 2.4. Differential gene expression analysis

Differential gene expression (DGE) analysis was performed on each dataset by Geo2R<sup>45</sup> to analyze the overexpression and down-regulation of important genes to distinguish between a healthy and diseased condition. GEO2R is an interactive web tool, provided by Gene Expression Omnibus (GEO) database, which is hosted by the National Center for Biotechnology Information (NCBI). GEO2R performs DGE analysis on RNA seq (the method is in beta version) and microarray data (more well-structured) by DESeq2 and limma, respectively. Thereby, during dataset selection, it was ensured that the study is conducted with microarray technology. GEO2R uses a variety of R packages from the Bioconductor project, which is an open-source software project based on the R programming language and provides tools for the analysis of high-throughput genomic data. In our case, we chose GEOquery and limma to perform differential expression analysis using original submitter-supplied processed data tables as input by GEO2R.

GEOquery parses GEO data into R data structures that can be used by other R packages. Limma (Linear Models for Microarray Analysis) is a statistical test for identifying differentially expressed genes in microarray data. It handles a wide range of experimental designs and data types and applies multiple-testing corrections to P-values to help correct for the occurrence of false-positives. In the tool, two groups were defined as healthy and cancerous groups. Benjamini & Hochberg's adjustment method was applied to the p-value to reduce false-positive results. Along with that, our limma analysis utilizes linear models to estimate gene expression differences while accounting for various sources of variability and noise in the data. It applies empirical Bayesian methods to borrow information across genes, which helps improve the estimation of fold changes and statistical significance. However, force normalization was applied on GSE46408, as the normalization method for the preprocessed data was unclear. The significance level cut-off was set to 0.05. The others are left as default. For the analysis, GEO2R used R programming language (version 4.2.2), Biobase (version 2.58.0), GEOquery (2.66.0), and Limma (version 3.54.0). Results are presented as a table of genes ordered by P-value, and as a collection of graphic plots to help visualize differentially expressed genes and assess data set quality.

Differentially expressed genes were extracted from each dataset, containing annotated genes and statistical values provided by GEO2R, using the “dplyr” package (version 1.0.6) in R (version 4.0.4). Significant genes were identified as having a p-value less than 0.05 and a Log<sub>2</sub>fold change value more than 1 ( $\log_2\text{fold} > 1$ ) for up-regulated genes or less than 1 ( $\log_2\text{fold} < 1$ ) for down-regulated genes. To gain a better understanding of the gene expression, volcano plots of selected up and down-regulated genes for each disease were constructed using Galaxy<sup>68</sup> (version 21.05.1) (an open-source, web-based platform for data-intensive biomedical research).

In Galaxy, the p-value for the plotting was set to less than 0.05 to display significantly upregulated and downregulated genes. The negative logarithm of the p-value (-log<sub>10</sub>P-value) was plotted against log<sub>2</sub>FC for each disease condition, providing clearer insights into the up and down-regulated genes, where log<sub>2</sub>FC  $> 1$  was labeled up-regulated and log<sub>2</sub>FC  $< 1$  was labeled down-regulated. To compare the filtered datasets for the cancers with T2DM, we loaded each dataset in the Venn Diagram functionality of FunRich<sup>69</sup> to have common genes between diseases. Once the diagram was plotted, genes in the intersected area were exported to an Excel formatted file. Derived common genes from FunRich were then modified into two tables of nodes, one of which contained the disease name and the other for the genes. Six such tables were loaded in Cytoscape<sup>48</sup> (version 3.8.2), respectively, to have a graphical representation of the disease-gene network.

## 2.5. Identification of molecular pathways

Differentially expressed genes (DEGs) were subjected to pathway enrichment analysis to identify the functional biological terminology and signal transduction pathways related to the disease phenotype. EnrichR<sup>46</sup> was utilized to conduct pathway enrichment analysis to figure out possible pathways that might be responsible for comorbidity between T2DM and cancer. All common genes (both up and down-regulated) were uploaded in the tool to shed light on the possible molecular mechanism. It provides a p-value, adjusted odds ratio, and combined score for each pathway along with the input genes that could be involved in that specific pathway. Thereby, significant pathways were selected based on the lowest p-value ( $P_{\text{Value}} < 0.01$ ), and the number of gene counts (how many genes were found to be involved in the pathway). Selected pathways were then studied for the mechanism of involvement in disease comorbidities through published research articles. Seven databases were taken into consideration to select the pathways: KEGG,<sup>70</sup> Bioplanet,<sup>71</sup> Reactome,<sup>72</sup> WikiPathway,<sup>73</sup> Elsevier Pathway Collection, Human Cyc,<sup>74</sup> and MSigDB Hallmark.<sup>75</sup>

## 2.6. Protein-protein interaction analysis

A protein-protein interaction (PPI) network was constructed to uncover potential links between DEG-encoded proteins found in both T2DM and cancers. Along with the dataset that includes all common genes (both up and down), individual disease-specific common genes were also taken into consideration for a specified analysis. The STRING database was utilized for PPI analysis using NetworkAnalyst.<sup>47</sup> A list of all common genes was uploaded in NetworkAnalyst<sup>47</sup> as an official gene symbol. A zero-degree network construction was made based on the criteria of degree  $\geq 25$  selected by the degree filter. The cut-off score for PPI analysis was defined as 900. Further, the obtained network was analyzed with cytoHubba<sup>76</sup> (version 0.1) in Cytoscape.<sup>48</sup> A hub protein network was constructed using cytoHubba, considering the top 18 proteins with the highest degree of interaction (based on topological parameters) with their closest neighbour proteins. The seed proteins (proteins coexisting in diseased datasets) with a degree of 8 or more were selected to locate the biomarkers that may be associated with each comorbidity. The functional descriptions of hub genes were validated by HUGO Gene Nomenclature Committee<sup>77</sup> and the NCBI Gene database.<sup>78</sup>

## 2.7. Cancer dependency analysis

Genetic dependency on cancer refers to the reliance of cancer cells on specific genetic alterations or mutations for their survival and proliferation. These alterations can occur in various genes that are involved in controlling cell growth, division, and death. A dependency study is conducted by inhibiting or knocking out the gene product or gene to see the impacts on cancer cell growth. The DepMap portal (<https://depmap.org/portal>), is a database that relies on the profiling of hundreds of cancer cell line models to provide analytical and visualization tools for investigating cancer dependencies. These databases utilized RNAi and CRISPR, two genome-wide silencing approaches to construct a complete map for cancer cells, contributed by scientists all over the world. RNAi knockdown is based on inhibition in gene expression at the mRNA level, while in CRISPR, the knockout directly works on gene.<sup>51,52</sup>

## 2.8. Protein level verification

We utilized the ULCAN (University of Alabama at Birmingham Cancer)<sup>49,50</sup> data analysis portal for analyzing the protein expression in specific cancers. Our analysis included data on four types of cancer, namely BRCA, HCC, CRC, and PRAD, however, we did not find any data regarding BLCA. Proteomic expression levels were verified for significant genes, and a graph was plotted to represent the expression in the z-score. The expression levels of proteins in 18 non-cancerous and 125 breast cancer samples were obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC).

## 2.9. Interaction analysis of transcription factors and miRNA with associated genes

We further analyzed the associated genes by introducing “NetworkAnalyst”<sup>47</sup> to focus on the precise role of transcription factors and miRNA in the regulatory pathway of gene expression. Thus, two networks were constructed to conclude the overall scenario. DEG regulatory trans-element was determined through the interaction between transcription factors and associated genes and was modelled using the JASPAR database.<sup>79</sup> The TF-gene interaction network was formed by considering the network topological parameters (i.e., degree and betweenness centrality) of identified transcription factors with degree value  $\geq 10$ . The top 16 TF were selected to centralize the regulation novelty. To describe the effect of regulatory miRNAs on DEGs at the posttranscriptional level, a miRNA and associated gene network was constructed in NetworkAnalyst<sup>47</sup> with a degree greater than or equal to 25 using miRTarbase.<sup>80</sup> The top five miRNA was selected based on the highest association with the regulated genes. To conclude the regulation by TF and miRNA, individual common gene datasets were also considered along with the dataset that represents all common genes regardless of up and down-regulation.

## 3. Result

### 3.1. Gene Expression analysis

The normalized gene expression was devoid of outliers, stating almost similar median and IQR for each sample within a dataset, visualized in the boxplot (Fig. S1). This indicates the relatively comparable expression levels between the two groups (healthy as control group and cancerous as test group). Subsequently, following log transformation and normalization procedures, the gene expression distributions were probed more comprehensively through density plots (Fig. S2) which also exhibit the analogous gene expression profiles between healthy and afflicted subjects. Further, to rigorously assess the statistical significance of differential gene expression, we employed t-statistics (Fig. S3). The near-linear alignment of these plots with the anticipated distribution, namely the student's t distribution, attests to the reliability and robustness of our statistical tests in discerning the significance of gene expression. This ensures the reliability of the statistical tests performed to assess the statistical significance of differential expression for each gene. Principal component analysis plots (Fig. S4) also show promising clustering, which helps us to proceed to further analysis.

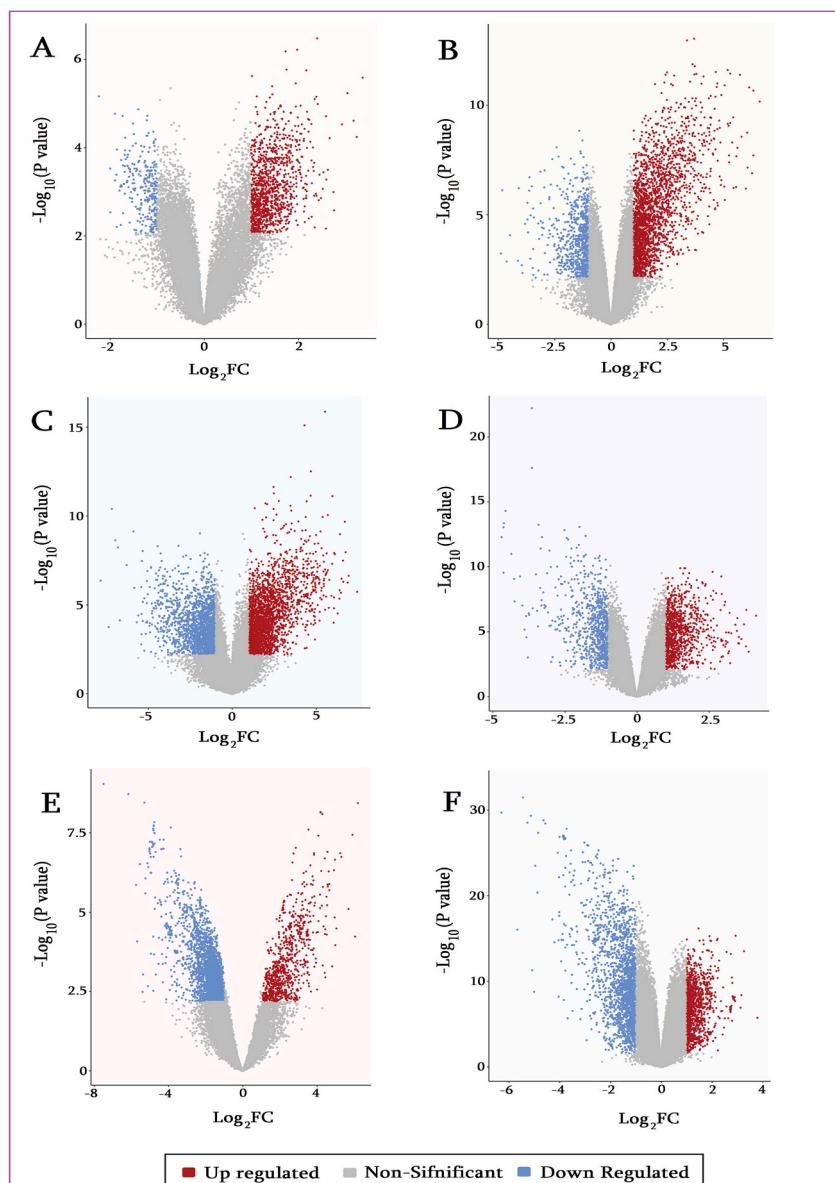
A total of 1226, 2098, 2938, 1310, 2585, and 1892 differentially expressed genes were found in the case of T2DM, BLCA, BRCA,

**Table 1**  
Datasets and resulting up and down-regulated genes.

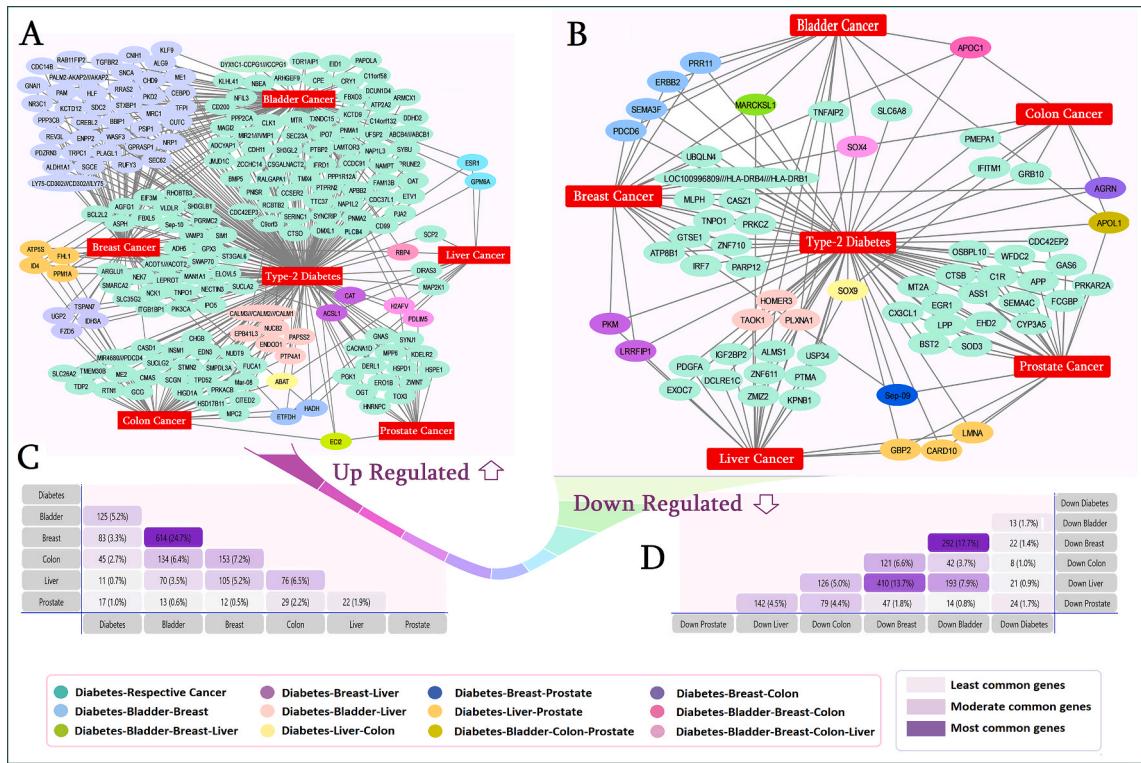
Disease Name	GEO Accession	GEO platform	Control Samples	Case Samples	Raw genes	Up-regulated genes	Down-regulated genes
Type-2 Diabetes	GSE25724	Affymetrix Human Genome U133A Array	7	6	22283	1017	209
Bladder Cancer	GSE7476	Affymetrix Human Genome U133 Plus 2.0 Array	3	9	54675	1514	584
Breast Cancer	GSE21422	Affymetrix Human Genome U133 Plus 2.0 Array	5	14	54675	1585	1354
Colon Cancer	GSE41328	Affymetrix Human Genome U133 Plus 2.0 Micro Array	10	10	54675	708	602
Liver Cancer	GSE46408	Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F	6	6	34958	545	2040
Prostate Cancer	GSE38241	Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F	21	18	45015	623	1269

CRC, HCC, and PRAD, respectively. For T2DM, 1017 up-regulated and 209 down-regulated genes were filtered from 22283 raw genes (Table 1) that originated as a result of differential gene expression analysis in Geo2R.<sup>45</sup> BRCA and BLCA had the highest number of up-regulated genes, with 1584 and 1514, respectively, while HCC was associated with the lowest number of up-regulated genes (545 genes). HCC was associated with 2040 down-regulated genes, followed by BRCA with 1354. Other malignancies, excluding BRCA and CRC, had more gene expression either in up-regulated or down-regulated genes. The quantity of expression of total genes, including up-regulated, down-regulated, and non-significant genes, are presented in the volcano plots (Fig. 1). By analyzing the differentially expressed genes, a total of 125, 83, 45, 11, and 17 commonly up-regulated genes (Fig. 2C) and 13, 22, 8, 21, and 24 down-regulated genes (Fig. 2D) were identified between T2DM and BLCA, BRCA, CRC, HCC, and PRAD, respectively. This co-expression percentage of T2DM with BLCA, BRCA, CRC, HCC, and PRAD turns into 5.2 %, 3.3 %, 2.7 %, 0.7 %, and 1 % upregulated genes (Figs. 2C) and 1.7 %, 1.4 %, 1 %, 0.9 %, 1.7 % downregulated genes (Fig. 2D). T2DM was observed to have the highest expressed shared genes with BLCA and BRCA. The case is justified in gene downregulation, where the percentage is 1.7 and 1.4 with BLCA and BRCA, respectively. However, with PRAD, only 1 % of common genes were up-regulated, while 1.7 % of common genes were downregulated.

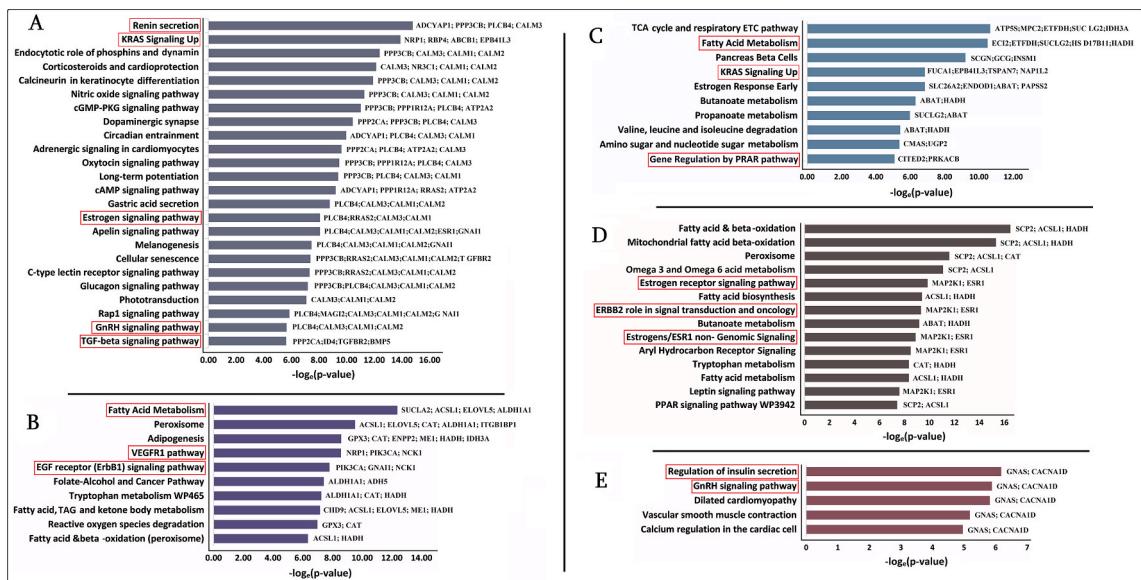
A clear visualization from the gene-disease network elucidated some particular genes having a higher degree of association with the comorbidities (Fig. 2A and B). Among those ATP5S, FHL1, ID4, PPM1A, ECI2, ETFDH, HADH, RBP4 are the most prominent in up-



**Fig. 1.** Volcano plots for each disease. The volcano plot depicts the abundance of up-regulated (red), down-regulated(blue), and non-significant genes (grey) for (A) Type-2 Diabetes, (B) Bladder Cancer, (C) Breast Cancer, (D) Colon Cancer, (E) Liver Cancer, and (F) Prostate Cancer.



**Fig. 2.** Disease-Gene network for dysregulated common genes. Colour grading in the diseased gene network implies the number of common (A) upregulated and (B) downregulated genes. Genes that are associated with at least three diseases are considered the most significant. The number and co-expression percentages of upregulated (C) and downregulated (D) common genes between the comorbidities are also shown in a subsequent chart. The number of common genes is represented using a colour gradient ranging from low (light purple) to high (dark purple). Note: Figure A and B were generated using Cytoscape<sup>48</sup> (version 3.8.2), and figure C and D were generated using FunRich.<sup>69</sup> All the figures were structured using Adobe Illustrator and Adobe Photoshop.



**Fig. 3.** Common pathways between type 2 diabetes and various cancers based on pathway enrichment analysis. Common pathways between type 2 diabetes and (A) Bladder cancer, (B) Breast cancer, (C) Colon cancer, (D) Liver cancer (E) Prostate Cancer has drawn in a bar diagram by plotting the pathway against the negative natural logarithm of the p-value. Significant cancer-related pathways, based on p-value, common genes, and hub proteins are denoted with red rectangles.

regulation scenario (Fig. 2A) where MRCKSL1, APOL1, APOC1, SOX4 draw attention with down-regulation (Fig. 2B). SOX4 was found to be downregulated in four cancers along with T2DM.

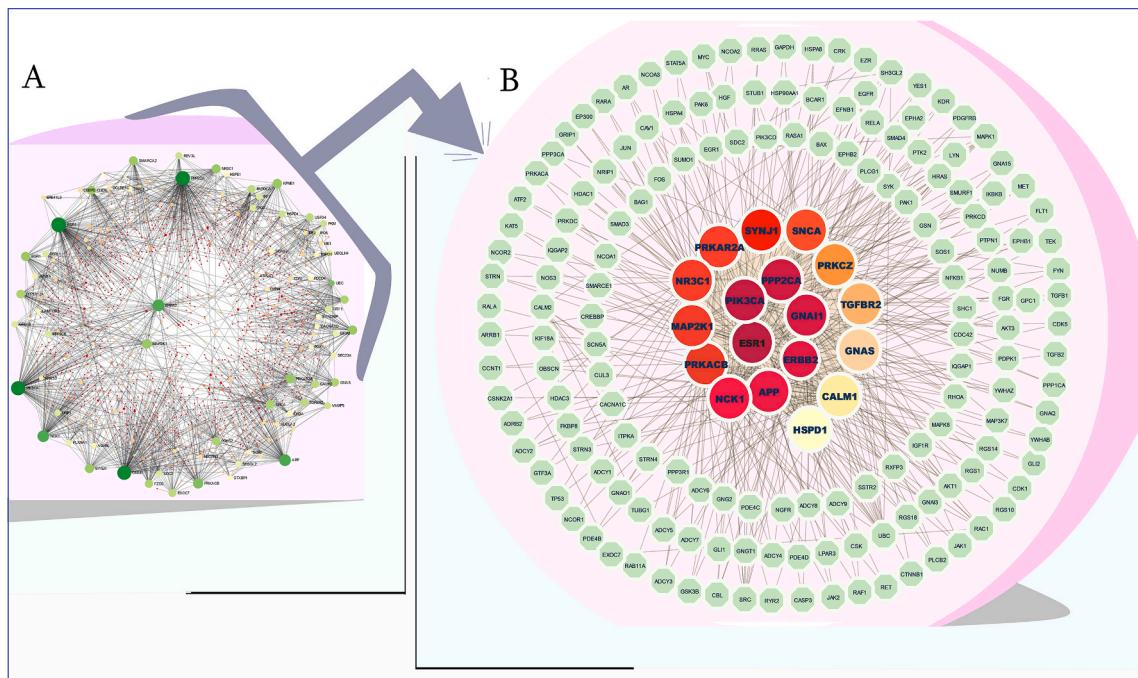
A comparison of these genes was made with other analyzed datasets derived from arrayexpress.<sup>61</sup> Our studied datasets (datasets from GEO) have some level of similarity with the compared datasets (datasets from arrayExpress) (Fig. S5). The most similarity is seen in BLCA and BRCA. Around 1368 genes of BLCA among 2099 genes (both in the context of up and down-regulation) are seen to be present in the compared dataset. This interesting scenario is also prevalent in other cases where the similarity holds for 1898, 744, 205, and 407 genes among 2938, 1310, 1892, and 1226 genes for BRCA, CRC, PRAD, and T2DM respectively with the compared datasets.

### 3.2. Analysis of associate pathway

Pathway enrichment analysis identified common associations between T2DM and selected carcinomas in estrogen signaling, fatty acid metabolism, and KRAS signaling pathways (Fig. 3). Based on the p-value, Renin secretion and the KRAS signaling pathway were considered significant in T2DM and BLCA comorbidity, however, Estrogen signalling and TGF $\beta$  signalling pathway may be significant since they share the key hub protein ESR1 (Fig. 3A). Fatty acid metabolism is integrally connected to each BRCA, CRC, and HCC association with type 2 diabetes. VEGFR1 pathway and EGF receptor (ErbB1) signalling pathway are essential for BRCA, whereas Estrogen and Leptin signalling pathways are crucial for T2DM and HCC comorbidity (Fig. 3D). Fatty acid metabolism has been identified as a common pathway between T2DM and BRCA comorbidity (Fig. 3B), while also potentially promoting colon cancer in diabetic individuals (Fig. 3C). We also identified common pathways, such as insulin secretion and GnRH signalling, between PRAD and T2DM, which may mechanistically link the two comorbidities (Fig. 3E). Our study also observed KRAS signalling to be involved in the comorbidity between T2DM and CRC, while GnRH signalling was found to be associated with the link between T2DM and BLCA. These common pathways, along with fatty acid metabolism and ESR1 signalling, may have roles in the interplay between T2DM and various carcinomas (Others are presented in Table S1).

### 3.3. Analysis of protein-protein interaction

Protein-protein interaction analysis guided the overview of associated dysregulated proteins that are interconnected with each other and contribute to any disease. This network was formed based on commonly expressed differential genes of the diseases with the STRING database via NetworkAnalyst<sup>47</sup> (Fig. 4A). A simplification of the protein-protein interaction resulted in hub protein visualization (Fig. 4B). A total of eighteen hub proteins were selected for further analysis based on their p-value and degree of association



**Fig. 4.** Depiction of protein-protein interaction network of dysregulated genes. The protein-protein interaction network of all dysregulated common genes between Type-2 Diabetes and various carcinomas (A). The protein's node size is ordered according to the degree of interaction with other proteins. B. Hub protein and its closest neighbour in interaction network of dysregulated common genes between Type-2 diabetes and Cancers. Significant protein is level with colour grading (red → orange). The neighbouring proteins are marked with a light-green colour that establishes the interconnection between the hub proteins. Note: Figure A was generated using NetworkAnalyst<sup>47</sup> and B using Cytoscape<sup>48</sup> (version 3.8.2). Finally, all the figures were structured using Adobe Illustrator and Adobe Photoshop.

(Table S2). The identified hub proteins are designated as ESR1, PIK3CA, PPP2CA, GNAI1, ERBB2, APP, NCK1, PRKACB, MAP2K1, NR3C1, PRKAR2A, SYNJ1, SNCA, PRKCZ, TGFBR2, GNAS, CALM1, HSPD1.

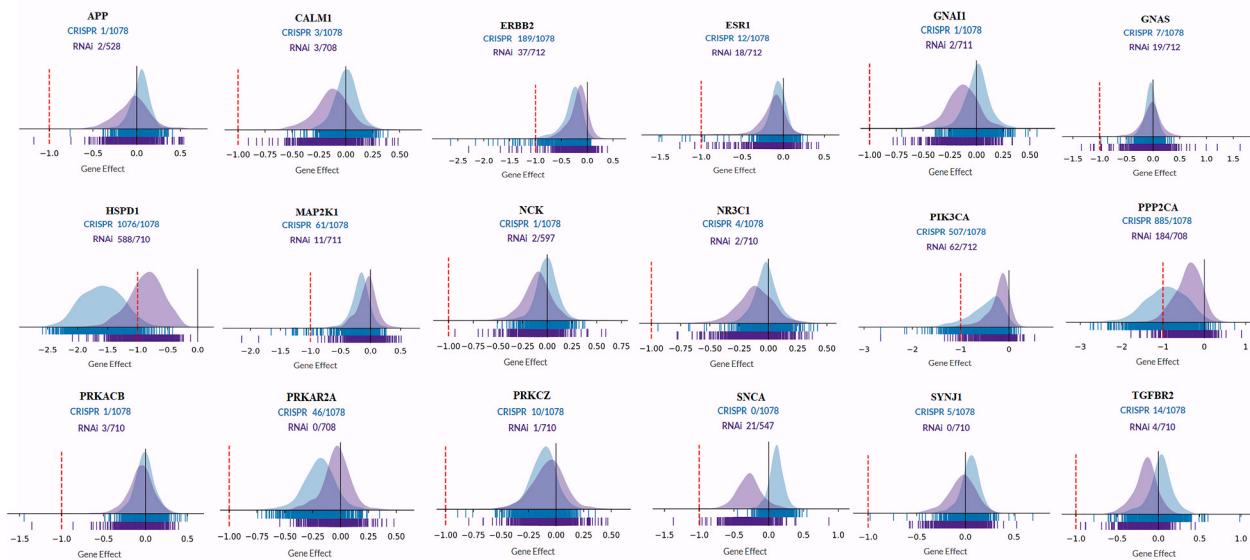
As observed all the 18 hub proteins were significant in the context of fold change and P value, associated with different comorbidities. Among the 18 hub proteins, all 4 proteins namely, ERBB2, APP, PRKAR2A, and PRKCZ were significant in cancer-T2DM comorbidities through down-regulation, whereas the remaining 14 showed significance with up-regulation (Table S2).

ESR1, PPP2CA, GNAI1, NR3C1, SNCA, TGFBR2, CALM1 namely, seven proteins were found to be associated with Type 2 diabetic and bladder cancer comorbidities, based on p-value, whereas PIK3CA, GNAI1, NCK1, NR3C1, SNCA, TGFBR2 linked Type 2 diabetic with breast cancer. PRKACB, CALM1, ESR1, and MAP2K1 are significant for CRC and HCC comorbidities with T2DM, while SYNJ1, GNAS, and HSPD1 are relevant for PRAD.

ESR1 shows a significant correlation with comorbidities of both bladder cancer (BLCA) and hepatocellular carcinoma (HCC), whereas CALM1 is responsible for the comorbidities associated with both BLCA and CRC. The genes GNAI1, NR3C1, SNCA, and TGFBR2 were found to be upregulated and have been associated with comorbidities of BLCA and BRCA in patients with T2DM, whereas ERBB2 was observed to be downregulated. While CRC and HCC showed no significant downregulated proteins, two of the hub proteins namely, ERBB2, and PRKCZ were downregulated in BRCA. While the down-regulated proteins in PRAD are largely unrelated to comorbidities with T2DM, six proteins, namely ESR1, APP, PRKAR2A, SNCA, TGFBR2, and CALM1, were identified as significant. Among these proteins, APP and PRKAR2A might be responsible for the co-morbid relationship between T2DM and cancer.

### 3.4. Cancer dependency

Cancer dependency refers to the reliance of cancer cells on specific genes or genetic alterations for their survival, growth, or other essential cellular functions. Cancer dependency analysis aims to identify and characterize these genetic dependencies in cancer. This analysis involves the study of gene essentiality or genetic vulnerabilities in cancer cells, where the loss or inhibition of specific genes or pathways leads to impaired cancer cell viability or growth. Genetic dependency studies are typically carried out using large-scale screening approaches, such as RNA interference (RNAi), CRISPR-Cas9 gene editing, or pharmacological inhibition. In our case, we presented the first two. Based on our analysis using the DepMap portal (Broad Institute of MIT and Harvard, 2021), we found that five members (ERBB2, ESR1, GNAS, MAP2K1, PIK3CA) were highly selective in both RNAi and CRISPR silencing, meaning that they were strongly associated with cancer cell dependencies (Fig. 5). Additionally, four members (NR3C1, PRKACB, SYNJ1, TGFBR2) were highly selective in CRISPR knockout only, indicating their association with cancer cell dependencies when using the CRISPR method for gene editing (Fig. 5). HSPD1 was commonly essential in both scenarios whereas PPP2CA was commonly essential in CRISPR knockout. The gene effect for associated cancer had been marked with tumour stage for the affected cell line (Fig. S6). Interestingly, five proteins (GNAI1, SNCA, TGFBR2, ERBB2, NR3C1) showed significance in BLCA and BRCA in terms of the primary and metastatic stage for designated cell lines. ESR1 had impacts on tumour progression for both BLCA and HCC while CALM1 might be involved in BLCA and CRC progression. All four genes (APP, SYNJ1, HSPD1, PRKAR2A) identified in PRAD progression were unique and affected



**Fig. 5.** Cancer dependency of significant Hub Proteins. Cancer dependency of top 18 HUB proteins in cancer cell lines, based on CRISPR and RNAi-mediated gene silencing data obtained from Depmap. The scale at the bottom of the graph indicates the effects of a particular gene on cancer progression. The scale at the bottom of the graph represents the effects of a particular gene on cancer progression, where a lower score indicates the gene's importance for a particular cell line. The vertical black line indicates a score of 0, meaning the gene/gene product is unnecessary for cancer progression, while the dashed red line represents the median of all essential genes.

comparatively lesser cell lines (Fig. S6).

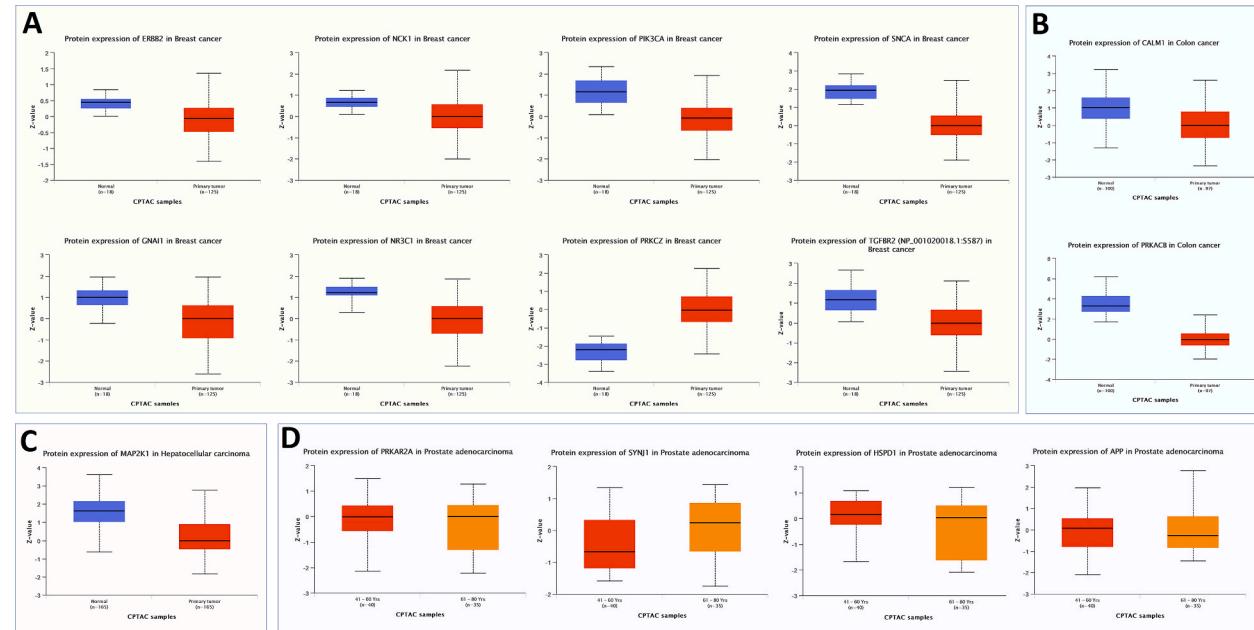
### 3.5. Protein level verification

The z-score-based data indicate the standard deviation from the median across the samples for various cancer types. The protein expression levels are determined by the population's area of expression for each protein. It should be noted that a p-value of less than 0.05 for a protein suggests significant expression during tumor progression. The expression of hub proteins in cancer and T2DM comorbidities were analyzed based on their z-scores, which indicate the deviation from the median expression level across samples for a specific cancer type (Fig. 6). The wider range of expression was observed for eight proteins, statistically significant, in BRCA namely ERBB2 (5.93E-03), NCK1(5.76E-07), PIK3CA (4.33E-08), SNCA (8.06E-14), GNAI1 (1.36E-08), NR3C1 (8.13E-11), PRKCZ (5.41E-16), and TGFBR2 (1.63E-03), while MAP2K1 (5.91E-31) had a wider range of expression in HCC. Conversely, CALM1 (2.63E-12) and PRKACB (3.63E-60) had wider expression in CRC, which was reflected in the higher standard deviation of their z-scores (Fig. 6). However, for PRAD, a limited variation in the expression of proteins was observed, including PRKAR2A (5.63E-01), SYNJ1 (2.01E-01), HSPD1 (3.61E-01), and APP (9.22E-01). These findings suggest that these proteins may play important roles in the pathogenesis and progression of these cancer types, and their differential expression may have clinical implications for cancer diagnosis, prognosis, and treatment.

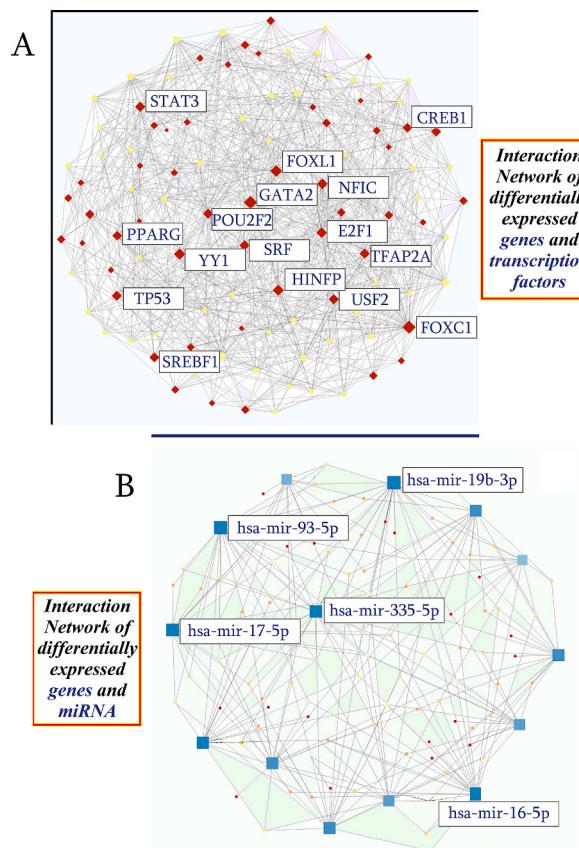
### 3.6. Transcription and miRNA interaction analysis

Transcription factors (TF) contribute to the initiation and regulation of gene transcription, whereas miRNAs are short RNA species that contribute to the posttranscriptional regulation of gene expression. Among the predicted *trans*-elements in the TF interaction network, we selected 16 transcription factors (GATA2, FOXC1, FOXL1, YY1, HINFP, NFIC, E2F1, STAT3, POU2F2, SRF, TFAP2A, USF2, CREB1, TP53, SREBF1, PPARG) with the highest degree of association with the common gene between T2DM and cancers (Fig. 7A). Of the 16 predicted transcription factors (TF), all promote gene expression in BRCA, while 14 out of 16 (GATA2, FOXC1, FOXL1, YY1, HINFP, NFIC, E2F1, STAT3, POU2F2, SRF, USF2, CREB1, TP53, PPARG) were related to the up-regulation of common genes between T2DM and BLCA. Additionally, 8 TF (USF2, CREB1, FOXC1, GATA2, YY1, NFIC, TFAP2A, POU2F2) were associated with the down-regulated genes for BLCA, while 12 TF were associated with the down-regulated genes for BRCA (GATA2, FOXC1, FOXL1, YY1, HINFP, NFIC, E2F1, SRF, TFAP2A, CREB1, TP53, PPARG).

The number of regulatory *trans*-elements decreased for common genes between T2DM and CRC, HCC, and PRAD. The predicted top miRNAs with the highest degree of association regulate all identified common genes between T2DM and BLCA, BRCA, CRC, HCC, and PRAD (Fig. 7B). A brief description of the functions of five specific miRNAs, namely hsa-mir-19b-3, hsa-mir-16-5p, hsa-mir-93-3p, hsa-mir-17-35, and hsa-mir-335-5p was provided (Table S3). In miRNA interaction analysis, hsa-mir-335-5p was discovered to regulate the majority of up and downregulated T2DM-associated disease genes. On the contrary, hsa-mir-16-5p was identified as a



**Fig. 6.** Protein expression of significant hub genes described by ULCAN Box Plot. The box plot summarizes the distribution of protein expression z-scores across cancer types, including (A) BRCA, (B) CRC, (C) HCC, and (D) PRAD. The z-score data represents the standard deviation (SD) from the median across samples for each cancer type.



**Fig. 7.** Interaction network of differentially expressed genes and regulatory elements. (A) Interaction network of differentially expressed genes and Transcription factors that regulate the genes. Square-shaped nodes indicate the transcription factors, while circle-shaped nodes indicate the proteins regulated by the transcription factors. (B) Interaction network of Differentially expressed genes and miRNA that regulate the genes. Square-shaped nodes indicate the miRNAs and circle-shaped nodes indicate the mRNA whose translation is being regulated by the miRNAs. Note: Figures were generated using NetworkAnalyst.<sup>47</sup> Finally, all the figures were structured using Adobe Illustrator and Adobe Photoshop.

regulatory factor in the up-regulation of genes in BLCA, BRCA, and PRAD and the down-regulation of CRC-associated genes. Also, the hsa-mir-93-5p controls the upregulation of BLCA and BRCA genes. Consequently, these three discovered miRNAs may contribute to the co-morbid pathophysiology of T2DM and selected malignancies.

#### 4. Discussion

Cancer, a persistent global health challenge, ranks among the leading causes of mortality worldwide, as documented by the World Health Organization (WHO). Environmental exposure, genetic predisposition, viral infection, and impaired immune function may contribute to cancer progression, while comorbidities such as diabetes mellitus influence its high mortality rate.<sup>81</sup> Individuals with type 2 diabetes have a 1.5-fold higher risk of cancer-related death for colorectal, pancreatic, liver, and endometrial cancer compared to the non-diabetic population.<sup>81</sup> In the present era, T2DM has been identified as one of the most prevalent diseases among various ethnic groups.

Previous clinical, epidemiological, and cross-sectional studies have found positive associations between T2DM and various forms of prevalent malignancies, including Bladder cancer (BLCA),<sup>37</sup> Breast cancer (BRCA),<sup>38,39</sup> Colon cancer (CRC),<sup>40,41</sup> Liver cancer (HCC),<sup>42</sup> and Prostate cancer (PRAD).<sup>43,44</sup> Epidemiologic research has identified correlations between T2DM and BLCA, in which male diabetic individuals had a higher mortality risk than non-diabetic patients.<sup>82</sup> Further, a meta-analysis concluded that diabetes individuals have a 23 % greater risk of BRCA, the second most prevalent cancer in women.<sup>83</sup> Previous research has also confirmed the link between T2DM-induced hyperinsulinemia and hyperglycemia and the increased risk of CRC.<sup>40,41</sup> The risk of CRC for type-2 diabetic individuals was determined to be 27 % higher by another study.<sup>84</sup> T2DM can also be identified as a risk factor for HCC as it strongly correlates with obesity.<sup>42,61</sup> Regarding PRAD, the most common carcinoma in men, patients with a history of diabetes had a 29 % increase in PRAD-specific fatality.<sup>85</sup> A recent study provided insights into the linkage between diabetes and tumour growth in BRCA in mice models. The study reported T2DM-induced hyperglycemia promoting glycation and cross-linking of the extracellular matrix of tumour tissue, increasing its malignancy.<sup>86</sup> Therefore, these studies provide evidence of the potential impact of diabetes on the pathogenesis and severity of cancer.

Apart from epidemiological, in-vitro, and in-vivo studies, integrated bioinformatics analysis has also provided insight into the association between Diabetes Mellitus and malignancies such as pancreatic cancer, hepatocellular carcinoma, and gastric Cancer.<sup>87–90</sup> One such study exploring the association study between T2DM and hepatocellular carcinoma (HCC), reported 98 DGE and ten hub proteins BUB1, CDCA8, DLGAP5, ASPM, POLQ, CENPE, WDHD1, HEL LS, TRIP13, and DEPDC1 which can be a responsible factor in diseases aetiology.<sup>88</sup> Another research identified genes of protein caspase 3 (CASP3) and tumour protein P53 (TP53) as core contributors to the pathogenesis of concurrent gastric cancers and T2DM.<sup>90</sup> In a separate study utilizing computational approaches, the authors identified 44 shared genes between T2DM and pancreatic cancers.<sup>87</sup> Subsequent experimental investigations revealed that one of these shared genes, S100A6, significantly promotes the prognosis of both diseases.

The existing research suggested that Type 2 diabetes has a complicated relationship with cancers involving multiple molecular interactions, which present significant challenges in treatments and hinder timely diagnosis.<sup>13</sup> Therefore, we aimed to analyze the co-expressed genes between T2DM and high-mortality cancers to uncover the underlying molecular interaction.<sup>89</sup> To ensure the robustness and validity of our findings, we executed rigorous statistical analysis on publicly available expression datasets used in our study. Our analysis revealed that the.

Following the preliminary analyses, we identified a total of 1226, 2099, 2938, 1310, 2585, and 1892 differentially expressed genes in T2DM, BLCA, BRCA, CRC, HCC, and PRAD, respectively. Most of these genes showed a similarity with the datasets we used for comparison (datasets from ArrayExpress). The up-regulated and down-regulated common genes were also identified between T2DM and selected malignancies (Fig. 2, C-D). Among the common genes, SOX4 was found to be downregulated in T2DM and BLCA, BRCA, HCC, and CRC. This SOX family protein is the most highly expressed in pancreatic islets<sup>91</sup> and regulates insulin secretion from pancreatic beta cells while also functioning in apoptosis and tumorigenesis.<sup>92</sup>

Followed by the differentially expressed genes, we attempted to explore the shared molecular pathway between cancer and T2DM using EnrichR<sup>46</sup> pathway enrichment analysis. Our investigation identified multiple pathways shared by T2DM and carcinomas, whose significance has been proven by prior studies. The renin secretion pathway was determined as the most significant in BLCA based on the p-value. This pathway also relies on cAMP and cGMP-PKG signalling.<sup>93</sup> Another renowned estrogen signalling pathway had been identified to link T2DM with cancers, especially BRCA. Based on supporting literature, a mechanism of this pathway could be driven by sex-hormone globulin. In T2DM, hyperinsulinemia lowers sex hormone-binding globulin,<sup>94</sup> resulting in an increase in the bioavailability of estrogen, which could lead to BRCA.<sup>95,96</sup> Our identified fat metabolism pathways are also mostly linked to the prognosis of BRCA, HCC, CRC, and T2DM (Fig. 3). Recent studies have shown a significant increase in total fatty acid levels during tumour progression and metastasis,<sup>97</sup> whereas elevated plasma fatty acid levels have also been documented in T2DM.<sup>98,99</sup> This could be an indication of a possible relationship between the comorbidities. Other significant pathways, represented in Fig. 3, are mostly involved in various cancer progression pathways and verified in the literature.

Further, we analyzed protein-protein interaction network to identify potential linkages between DEG-encoded proteins in T2DM and cancers (Fig. 4). The focus is drawn to 18 hub proteins, namely ESR1, PIK3CA, PPP2CA, GNAI1, ERBB2, APP, NCK1, PRKACB, MAP2K1, NR3C1, PRKAR2A, SYNJ1, SNCA, PRKCZ, TGFB2, GNAS, CALM1, HSPD1 by cytohubba<sup>76</sup> plugin of Cytoscape<sup>48</sup> (Fig. 4). Among all the identified hub proteins, we observed, ESR1, as the most significant protein linking T2DM and cancer, specifically BRCA, BLCA, and PRAD comorbidities (Table S2). According to previous research, ESR1 regulates the expression of estrogen-responsive genes involved in numerous physiological processes, including growth, metabolism, and reproductive functions. ESR1 was upregulated in both BLCA and BRCA while showing downregulation in PRAD (Table S2). In different cell lines of hepatocellular carcinomas, ESR1 also contributes to the metastatic stage (Fig. S6). Not to mention that its levels are also found to be elevated in individuals with type-2 diabetes.<sup>100</sup> Therefore, these proteins may be responsible for cancer pathogenesis in diabetic conditions. Our pathway enrichment analysis also identified a remarkable significance involving estrogen receptors (Fig. 3). In population studies, the role of ESR1 polymorphism in breast cancer development is well-studied.<sup>101</sup> Furthermore, it has been identified in the developing mechanism of T2DM,<sup>102</sup> HCC<sup>103</sup> and BLCA.<sup>104</sup> Therefore, it could potentially contribute to the comorbidities between T2DM and cancers.

Another significant protein, predicted by the degree of association (Table S2), is Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), a renowned proto-oncogene. PIK3CA is found to be elevated in both T2DM and BRCA, suggesting a co-morbid relationship as the imbalance of PIK3CA signalling in obesity and diabetes is well studied,<sup>105</sup> and mutation in this gene is responsible for breast and uterine cancer.<sup>106</sup> ERBB2, on the other hand, shows reduced expression in T2DM, BLCA, and BRCA and is statistically significant, characterized by p Value (Table S2). However, molecular studies found ERBB2 overexpressed in BRCA,<sup>107</sup> BLCA,<sup>108</sup> and other carcinomas. Therefore, there may be a mechanism in play that altered the expression of ERBB2 in T2DM comorbidities. The expression pattern of the remaining hub protein in different cancer and T2DM comorbidities was observed (Table S2). Most of the identified hub proteins functioned in DNA repair, cell cycle, growth regulation, cell division, and degradation through kinase and other signaling pathways (Table S4). The functionality is validated by HUGO Gene Nomenclature Committee<sup>77</sup> and NCBI gene database.<sup>78</sup>

In addition, we performed a miRNA-diseased gene interaction analysis to determine the regulation of the identified DEGs. Through the analysis, we identified several critical miRNAs that could be involved in the regulatory pathways (Fig. 7B). These miRNAs can be opted for cancer progression (Table S3) in a dysregulated manner. Our research predicted that hsa-mir-335-5p plays a crucial regulatory function in the overexpression and downregulation of most T2DM-associated disease genes. These functionalities can also contribute to prevailing BLCA, CRC, and HCC as supported by prior studies.<sup>109</sup> mir-335-5p was also discovered to suppress gastric cancer<sup>110</sup> while hsa-mir-16-5p contributes to the posttranscriptional regulation of genes in BLCA, BRCA, and PRAD (Fig. 7B), as well as playing a key regulatory function in the lung, brain, and cervical carcinomas by modulating cellular proliferation and apoptosis.<sup>111</sup> Furthermore, overexpression of mir-16-5p was also reported to prevent carcinogenic effects by inhibiting several oncogenes.<sup>112</sup> Among the other identified miRNA, Hsa-mir-93-5p may have regulatory roles in BLCA and BRCA carcinogenesis, while mir-19b-3p

was reported to promote CRC proliferation.<sup>113</sup>

The regulatory role of transcription factors (TF) in T2DM-cancer comorbidity was also evaluated in our study (Table S5). We identified 16 *trans*-elements responsible for the transcriptional regulation of commonly expressed diseased genes. Nearly every identified TF plays a crucial role by regulating the expression of the shared genes among T2DM and BLCA, BRCA, CRC, HCC, and PRAD (Fig. 7A). However, the highest number of transcription factor associations is observed in the transcriptional regulation of T2DM, BRCA, and BLCA-associated genes. Among the predicted TF, GATA2 has the highest degree factor, which regulates hematopoietic and endocrine cell development.<sup>114</sup> Moreover, FOXL1 has also been identified in prior studies to play key roles in the control of many processes during ontogenesis, such as gene expression, metabolism, and cell division, while also being reported to inhibit gastric and pancreatic tumours.<sup>115</sup> On the contrary, E2F1 regulates the cell cycle and functions with tumour suppressor proteins<sup>116</sup> while also regulating key metabolic processes.<sup>117</sup> Peroxisome proliferator-activated receptor gamma (PPARG) has been associated with the development of various disorders, such as cancers, obesity, diabetes, and atherosclerosis,<sup>118</sup> and might be used as a potential therapeutic target for type 2 diabetes and cancers.<sup>119</sup>

In redundancy analysis, these proteins appeared to affect cancer progression significantly, as dysregulation of those is marked noticeably in cancerous cells (Fig. 5). The specific effect of the hub genes in different organ systems and their involvement in primary or metastatic tumours are shown (Fig. S6). The significance of protein level was also determined in normal and cancer cells through the Z-score (Fig. 6), in which 15 hub protein expressions were confirmed in BRCA, HCC, CRC, and PRAD. We could not obtain experimental data for BLCA because of data unavailability.

Taken together, the study provides valuable insights into the intricate relationship between type 2 diabetes mellitus (T2DM) and high-mortality cancers. Although further experimental validation is needed, our findings offer promising avenues for future research. One significant discovery is the identification of the common gene Sox4, which may serve as a genetic marker linking T2DM and multiple cancers. Additionally, several hub proteins, including ESR1, PIK3CA, GNAI1, ERBB2, NR3C1, SNCA, and TGFBR2, have been validated as statistically significant biomarkers for cancer diagnosis in individuals with diabetes. This study highlights the importance of miRNAs, such as hsa-mir-335-5p, hsa-mir-16-5p, and hsa-mir-93-5p, along with the identified transcription factors, in modulating the expression of disease-associated genes. This opens up possibilities for utilizing these regulatory molecules to better understand and potentially intervene in the T2DM-cancer relationship. Notably, targeting significant pathways like fatty acid metabolism and the estrogen signalling pathway holds promise for the development of novel therapeutics for cancer treatment in individuals with T2DM.

Overall, this study provides a foundation for identifying novel biomarkers based on the association between T2DM and prevalent malignancies, emphasizing the need for further validation of the identified genes, proteins, miRNAs, and transcription factors as valuable tools for cancer diagnosis. By expanding our knowledge of these molecular mechanisms, we aim to inspire further research and contribute to improved strategies for the diagnosis, treatment, and management of individuals with both T2DM and cancer.

## 5. Conclusion

Our study pinpointed key genes and biomolecules underlying the co-occurrence of T2DM and prevalent carcinomas, spotlighting 18 key proteins, including ESR1, PIK3CA, GNAI1, ERBB2, NR3C1, SNCA, and TGFBR2, emerging as highly significant. Additionally, we identified 16 regulatory *trans*-elements and five pivotal miRNAs including hsa-mir-335-5p, hsa-mir-16-5p, and hsa-mir-93-5p that play pivotal roles in post-transcriptional gene regulation, particularly in T2DM, BLCA, and BRCA. Remarkably, our research highlights a significant comorbid connection between T2DM, BLCA, and BRCA, and a moderate one with PRAD. These findings offer potential biomarkers and valuable insights, while further translational research is essential for the development of precision cancer medicine.

### Data availability

The authors declare that all the data along with the codes, will be available without any restrictions. Please contact Dr Ajit Ghosh (Email: [aghosh-bmb@sust.edu](mailto:aghosh-bmb@sust.edu)) to request relevant data.

### Ethics approval and consent to participate

Not Applicable.

### Consent for publication

Not Applicable.

### Funding

This study was done with no external funding scenario.

### CRediT authorship contribution statement

**Abdullah Al Marzan:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Shatila Shahi:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Md**

**Sakil Arman:** Formal analysis, Investigation, Methodology, Writing – original draft. **Md Zafrul Hasan:** Supervision, Writing – review & editing. **Ajit Ghosh:** Conceptualization, Funding acquisition, Resources, Supervision, Validation, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

The authors acknowledge the NCBI GEO database and GEO2R for their platform and the researchers for uploading their valuable datasets and analysis. ChatGPT 3 (Mar 23 Version) was used for minimal language editing and to ensure grammatical accuracy.

### Abbreviations

T2DM	Type-2 Diabetes Mellitus
BLCA	Bladder Cancer
BRCA	Breast Cancer
CRC	Colorectal Cancer
HCC	Hepatocellular Carcinoma
PRAD	Prostate Adenocarcinoma
miRNAs	MicroRNAs
TF	Transcription Factor

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.abst.2023.10.001>.

### References

- 1 Saeedi P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas. In: *Diabetes Res Clin Pract*. ninth ed. vol. 157. 2019.
- 2 Galicia-Garcia U, et al. Pathophysiology of type 2 diabetes mellitus. *Int J Mol Sci.* 2020;21:6275.
- 3 Stumvoll M, Goldstein BJ, Van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet.* 2005;365:1333–1346.
- 4 Weyer C, Bogardus C, Mott DM, Pratley RE. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest.* 1999;104:787.
- 5 Barone BB. Long-term all-cause mortality in cancer patients with preexisting diabetes mellitus. *JAMA.* 2008;300:2754.
- 6 Ferlay J, E. M. L. F. C. M. M. L. P. M. Z. A. S. I. B. F. Global Cancer Observatory. *Cancer Today.* Lyon, France: International Agency for Research on Cancer; 2020.
- 7 Ferlay J, et al. Cancer statistics for the year 2020: an overview. *Int J Cancer.* 2021;149:778–789.
- 8 Sung H, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021; 71:209–249.
- 9 Greenwood M, Wood F. The relation between the cancer and diabetes death-rates. *J Hyg.* 1914;14:83–118.
- 10 Ling S, et al. Inequalities in cancer mortality trends in people with type 2 diabetes: 20 year population-based study in England. *Diabetologia.* 2023;66:657–673.
- 11 Setiawan VW, et al. Pancreatic cancer following incident diabetes in African Americans and Latinos: the multiethnic cohort. *JNCI J Natl Cancer Inst.* 2019;111: 27–33.
- 12 Shen B, et al. Association between age at diabetes onset or diabetes duration and subsequent risk of pancreatic cancer: results from a longitudinal cohort and mendelian randomization study. *Lancet Reg Health West Pac.* 2023;30, 100596.
- 13 Renzi C, et al. Comorbid chronic diseases and cancer diagnosis: disease-specific effects and underlying mechanisms. *Nat Rev Clin Oncol.* 2019;16:746–761.
- 14 Basu SK, Subratagoswami editors C. *Cancer Diagnostics and Therapeutics.* Springer Singapore; 2022. <https://doi.org/10.1007/978-981-16-4752-9>.
- 15 Klein AP. Genetic susceptibility to pancreatic cancer. *Mol Carcinog.* 2012;51:14–24.
- 16 Byers T, Sedjo RL. Body fatness as a cause of cancer: epidemiologic clues to biologic mechanisms. *Endocr Relat Cancer.* 2015;22:R125–R134.
- 17 Xu J, et al. Association between markers of glucose metabolism and risk of colorectal cancer. *BMJ Open.* 2016;6, e011430.
- 18 Stolzenberg-Solomon, Insulin RZ. Glucose, insulin resistance, and pancreatic cancer in male smokers. *JAMA.* 2005;294:2872.
- 19 Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet.* 2001;357:539–545.
- 20 Byers T, Sedjo RL. Body fatness as a cause of cancer: epidemiologic clues to biologic mechanisms. *Endocr Relat Cancer.* 2015;22:R125–R134.
- 21 Fuster JJ, et al. Noncanonical wnt signaling promotes obesity-induced adipose tissue inflammation and metabolic dysfunction independent of adipose tissue expansion. *Diabetes.* 2015;64:1235–1248.
- 22 Pate KT, et al. Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J.* 2014;33:1454–1473.
- 23 Donati G, et al. Epidermal Wnt/β-catenin signaling regulates adipocyte differentiation via secretion of adipogenic factors. *Proc Natl Acad Sci USA.* 2014;111: 3129–3139.
- 24 Tudzarova S, Osman MA. The double trouble of metabolic diseases: the diabetes–cancer link. *Mol Biol Cell.* 2015;26:3129–3139.
- 25 Buckbinder L, et al. Induction of the growth inhibitor IGF-binding protein 3 by p53. *Nature.* 1995. 1995;377:646–649, 6550 377.
- 26 Alfaro-Arnedo E, et al. IGF1R acts as a cancer-promoting factor in the tumor microenvironment facilitating lung metastasis implantation and progression. *Oncogene.* 2022. 2022;41:3625–3639, 28 41.
- 27 Wolf I, Sadetzki S, Catane R, Karasik A, Kaufman B. Diabetes mellitus and breast cancer. *Lancet Oncol.* 2005;6:103–111.
- 28 de Beer JC, Liebenberg L. Does cancer risk increase with HbA1c, independent of diabetes? *Br J Cancer.* 2014;110:2361–2368.
- 29 Warburg O. On the origin of cancer cells. *Science.* 1956;123:309–314.

- 30 Wu S, et al. Hyperglycemia enhances immunosuppression and aerobic glycolysis of pancreatic cancer through upregulating bni1-UPF1-HK2 pathway. *Cell Mol Gastroenterol Hepatol.* 2022;14:1146–1165.
- 31 Yu T, Robotham JL, Yoon Y. Increased production of reactive oxygen species in hyperglycemic conditions requires dynamic change of mitochondrial morphology. *Proc Natl Acad Sci U S A.* 2006;103:2653–2658.
- 32 Aggeli IK, Theofilatos D, Beis I, Gaitanaki C. Insulin-induced oxidative stress up-regulates heme oxygenase-1 via diverse signaling cascades in the C2 skeletal myoblast cell line. *Endocrinology.* 2011;152:1274–1283.
- 33 Valko M, Rhodes CJ, Moncol J, Izakovic M, Mazur M. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact.* 2006;160:1–40.
- 34 Lin W-W, Karin M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J Clin Invest.* 2007;117:1175–1183.
- 35 Valko M, et al. Free radicals and antioxidants in normal physiological functions and human disease. *Int J Biochem Cell Biol.* 2007;39:44–84.
- 36 Aggarwal BB. Nuclear factor- $\kappa$ B. *Cancer Cell.* 2004;6:203–208.
- 37 Huang WL, et al. Effect of diabetes mellitus and glycemic control on the prognosis of non-muscle invasive bladder cancer: a retrospective study. *BMC Urol.* 2020;20:1–7.
- 38 Xiong F, Wang J, Nierenberg JL, Schmajuk G, Graff RE. 1072-P: diabetes mellitus and risk of breast cancer: a large-scale, prospective, population-based cohort study. *Diabetes.* 2022;71.
- 39 Redaniel MTM, Jeffreys M, May MT, Ben-Shlomo Y, Martin RM. Associations of type 2 diabetes and diabetes treatment with breast cancer risk and mortality: a population-based cohort study among British women. *Cancer Causes Control.* 2012;23:1785–1795.
- 40 Ottaiano A, et al. Metastatic colorectal cancer and type 2 diabetes: prognostic and genetic interactions. *Mol Oncol.* 2022;16:319–332.
- 41 Li Z, et al. Type 2 diabetes and risk of early-onset colorectal cancer. *Gastro Hep Advances.* 2022;1:186–193.
- 42 Mao D, et al. Risk associations of glycemic burden and obesity with liver cancer—a 10-year analysis of 15,280 patients with type 2 diabetes. *Hepatol Commun.* 2022;6:1350–1360.
- 43 Bansal D, Bhansali A, Kapil G, Undela K, Tiwari P. Type 2 diabetes and risk of prostate cancer: a meta-analysis of observational studies. *Prostate Cancer and Prostatic Diseases.* 2013;2012;16:151–158, 2 16.
- 44 Sánchez-Maldonado JM, et al. Type 2 diabetes-related variants influence the risk of developing prostate cancer: a population-based case-control study and meta-analysis. *Cancers.* 2022;14:2376.
- 45 Barrett T, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2012;41:D991–D995.
- 46 Kuleshov MV, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44:W90–W97.
- 47 Xia J, Benner MJ, Hancock REW. NetworkAnalyst - integrative approaches for protein–protein interaction network analysis and visual exploration. *Nucleic Acids Res.* 2014;42:W167–W174.
- 48 Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–2504.
- 49 Chandrashekhar DS, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia.* 2017;19:649–658.
- 50 Chandrashekhar DS, et al. UALCAN: an update to the integrated cancer data analysis platform. *Neoplasia.* 2022;25:18–27.
- 51 Dwane L, et al. Project Score database: a resource for investigating cancer cell dependencies and prioritizing therapeutic targets. *Nucleic Acids Res.* 2021;49:D1365–D1372.
- 52 Tsherniak A, et al. Defining a cancer dependency map. *Cell.* 2017;170:564–576, e16.
- 53 Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol.* 2012;6:140.
- 54 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30:207–210.
- 55 Dominguez V, et al. Class II phosphoinositide 3-kinase regulates exocytosis of insulin granules in pancreatic beta cells. *J Biol Chem.* 2011;286:4216–4225.
- 56 Mengual L, et al. DNA microarray expression profiling of bladder cancer allows identification of noninvasive diagnostic markers. *J Urol.* 2009;182:741–748.
- 57 Kretschmer C, Conradi A, Kemmner W, Sterner-Kock A. Latent transforming growth factor binding protein 4 (LTBP4) is downregulated in mouse and human DCIS and mammary carcinomas. *Cell Oncol.* 2011;34:419–434.
- 58 Lin G, et al. Reproducibility Probability Score—incorporating measurement variability across laboratories for gene selection. *Nat Biotechnol.* 2006;24:1476–1477.
- 59 Chen Y-L, Wang T-H, Hsu H-C, Yuan R-H, Jeng Y-M. Overexpression of CTHRC1 in hepatocellular carcinoma promotes tumor invasion and predicts poor prognosis. *PLoS One.* 2013;8, e70324.
- 60 Aryee MJ, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med.* 2013;5:169, 10.
- 61 Parkinson H, et al. ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 2007;35:D747–D750.
- 62 Zhang Z, Furge KA, Yang XJ, Teh BT, Hansel DE. Comparative gene expression profiling analysis of urothelial carcinoma of the renal pelvis and bladder. *BMC Med Genom.* 2010;3:58.
- 63 Turashvili G, et al. Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer.* 2007;7:55.
- 64 Pesson M, et al. A gene expression and pre-mRNA splicing signature that marks the adenoma-adenocarcinoma progression in colorectal cancer. *PLoS One.* 2014;9, e87761.
- 65 Liao Y-L, et al. Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. *Oncogene.* 2008;27:5578–5589.
- 66 Marselli L, et al. Gene expression profiles of beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS One.* 2010;5, e11499.
- 67 Kim J, et al. SOD3 acts as a tumor suppressor in PC-3 prostate cancer cells via hydrogen peroxide accumulation. *Anticancer Res.* 2014;34:2821–2831.
- 68 Afgan E, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018;46:W537–W544.
- 69 Pathan M, et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics.* 2015;15:2597–2601.
- 70 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40:D109–D114.
- 71 Huang R, et al. The NCATS BioPlanet – an integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems Biology, and chemical genomics. *Front Pharmacol.* 2019;10.
- 72 Croft D, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 2011;39:D691–D697.
- 73 Pico AR, et al. WikiPathways: pathway editing for the people. *PLoS Biol.* 2008;6:e184.
- 74 Trupp, M. et al. Beyond the Genome (BTG) Is a (PGDB) Pathway Genome Database.
- 75 Liberzon A, et al. The molecular signatures database Hallmark gene set collection. *Cell Syst.* 2015;1:417–425.
- 76 Chin CH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8:1–7.
- 77 Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* 2021;49:D939–D946.
- 78 Brown GR, et al. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* 2015;43:D36–D42.
- 79 Khan A, 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.
- 80 Hsu S Da, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2011;39.
- 81 Ling S, et al. Inequalities in cancer mortality trends in people with type 2 diabetes: 20 year population-based study in England. *Diabetologia.* 2023;66:657–673.
- 82 Xu Y, Huo R, Chen X, Yu X. Diabetes mellitus and the risk of bladder cancer: a PRISMA-compliant meta-analysis of cohort studies. *Medicine.* 2017;96.
- 83 Larsson SC, Mantzoros CS, Wolk A. Diabetes mellitus and risk of breast cancer: a meta-analysis. *Int J Cancer.* 2007;121:856–862.

- 84 González N, et al. 2017 update on the relationship between diabetes and colorectal cancer: epidemiology, potential molecular mechanisms and therapeutic implications. *Oncotarget*. 2017;8, 18456.
- 85 Lee J, Giovannucci E, Jeon JY. Diabetes and mortality in patients with prostate cancer: a meta-analysis. *SpringerPlus*. 2016;5:1–11.
- 86 Wang W, et al. Diabetic hyperglycemia promotes primary tumor progression through glycation-induced tumor extracellular matrix stiffening. *Sci Adv*. 2022;8, eab01673.
- 87 Hu Y, et al. Identification of the shared gene signatures and biological mechanism in type 2 diabetes and pancreatic cancer. *Front Endocrinol*. 2022;13.
- 88 Maddah R, et al. Identification of critical genes and pathways associated with hepatocellular carcinoma and type 2 diabetes mellitus using integrated bioinformatics analysis. *Inform Med Unlocked*. 2022;30.
- 89 Li X, et al. Identification of key genes and pathways in peripheral blood mononuclear cells of type 1 diabetes mellitus by integrated bioinformatics analysis. *Diabetes Metab J*. 2022;46:451–463.
- 90 Liu S, et al. Identification of core genes associated with type 2 diabetes mellitus and gastric cancer by bioinformatics analysis. *Ann Transl Med*. 2022;10, 247–247.
- 91 Volkov P, et al. Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. *Diabetes*. 2017;66:1074–1085.
- 92 Hui W, et al. SOX4 overexpression regulates the p53-mediated apoptosis in hepatocellular carcinoma: clinical implication and functional analysis in vitro. *Carcinogenesis*. 2010;31:1298–1307.
- 93 Schramm A, et al. Protein kinase G is involved in acute but not in long-term regulation of renin secretion. *Front Pharmacol*. 2019;10.
- 94 Strain G, Zumoff B, Rosner W, Pi-Sunyer X. The relationship between serum levels of insulin and sex hormone-binding globulin in men: the effect of weight loss. *J Clin Endocrinol Metab*. 1994;79:1173–1176.
- 95 Dimou NL, et al. Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study. *Int J Epidemiol*. 2019;48:807–816.
- 96 Fortunati N, Catalano MG, Bocuzzi G, Frairia R. Sex Hormone-Binding Globulin (SHBG), estradiol and breast cancer. *Mol Cell Endocrinol*. 2010;316:86–92.
- 97 Xu S, et al. Fatty acid synthase promotes breast cancer metastasis by mediating changes in fatty acid metabolism. *Oncol Lett*. 2020;21, 1–1.
- 98 Boden G, Laakso M. Lipids and glucose in type 2 diabetes. *Diabetes Care*. 2004;27:2253–2259.
- 99 Ravikumar B, et al. Real-time assessment of postprandial fat storage in liver and skeletal muscle in health and type 2 diabetes. *Am J Physiol Endocrinol Metabol*. 2005;288:E789–E797.
- 100 Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*. 2004;4:579–591.
- 101 Yang W, et al. Impact of ESR1 polymorphisms on risk of breast cancer in the Chinese han population. *Clin Breast Cancer*. 2021;21:e235–e242.
- 102 Gregorio KCR, Laurindo CP, Machado UF. Estrogen and glycemic homeostasis: the fundamental role of nuclear estrogen receptors ESR1/ESR2 in glucose transporter GLUT4 regulation. *Cells*. 2021;10:99.
- 103 Meng J, et al. CCL23 suppresses liver cancer progression through the CCR1/AKT/ESR1 feedback loop. *Cancer Sci*. 2021;112:3099–3110.
- 104 Hsu I, Vitkus S, Da J, Yeh S. Role of oestrogen receptors in bladder cancer development. *Nat Rev Urol*. 2013;10:317–326.
- 105 Huang X, Liu G, Guo J, Su Z. The PI3K/AKT pathway in obesity and type 2 diabetes. *Int J Biol Sci*. 2018;14:1483–1496.
- 106 Campbell IG, et al. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res*. 2004;64:7678–7681.
- 107 Slamon DJ, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*. 1989;244:707–712.
- 108 Sauter G, Moch H, Gasser TC, Mihatsch MJ, Waldman FM. Heterogeneity of chromosome 17 and erbB-2 gene copy number in primary and metastatic bladder cancer. *Cytometry*. 1995;21:40–46.
- 109 Ye L, et al. Functions and targets of miR-335 in cancer. *OncoTargets Ther*. 2021;14:3335–3349.
- 110 Gao Y, et al. miR-335-5p suppresses gastric cancer progression by targeting MAPK10. *Cancer Cell Int*. 2021;21:1–12.
- 111 Ghafouri-Fard S, et al. A review on the role of mir-16-5p in the carcinogenesis. *Cancer Cell Int*. 2022;22:1–11.
- 112 Chava S, et al. miR-15a-5p, miR-15b-5p, and miR-16-5p inhibit tumor progression by directly targeting MYCN in neuroblastoma. *Mol Oncol*. 2020;14:180–196.
- 113 Jiang T, et al. miR-19b-3p promotes colon cancer proliferation and oxaliplatin-based chemoresistance by targeting SMAD4: validation by bioinformatics and experimental analyses. *J Exp Clin Cancer Res*. 2017;36:131.
- 114 Vicente C, Conchillo A, García-Sánchez MA, Odero MD. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit Rev Oncol Hematol*. 2012;82:1–17.
- 115 Zhang G, et al. FOXL1, a novel candidate tumor suppressor, inhibits tumor aggressiveness and predicts outcome in human pancreatic cancer. *Cancer Res*. 2013; 73:5416–5425.
- 116 Pan H, et al. Key roles for E2F1 in signaling p53-dependent apoptosis and in cell division within developing tumors. *Mol Cell*. 1998;2:283–292.
- 117 Denechaud PD, Fajas L, Giralt A. E2F1, a novel regulator of metabolism. *Front Endocrinol*. 2017;8:311.
- 118 Peters JM, Shah YM, Gonzalez FJ. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nat Rev Cancer*. 2012;12:181.
- 119 Frkic RL, Richter K, Bruning JB. The therapeutic potential of inhibiting PPAR $\gamma$  phosphorylation to treat type 2 diabetes. *J Biol Chem*. 2021;297, 101030. HumanCyc. *Genome Biol* 11, O12 (2010).