ACCSCIENCE PUBLISHING

Tumor Discovery

ORIGINAL RESEARCH ARTICLE

Systemic drug repurposing for pancreatic cancer based on genetic and epigenetic network analysis using a systems biology approach and deep neural learning of drug-target interactions

Yi-Hsin Tsai and Bor-Sen Chen*

Laboratory of Automatic Control, Signal Processing and Systems Biology, Department of Electrical Engineering, Institute of Electronic Engineering, National Tsing Hua University, Hsinchu, Taiwan, China

Abstract

Pancreatic cancer is a malignant tumor associated with a high mortality rate. This research presents a systems biology approach to explore the mechanisms of pancreatic ductal adenocarcinoma (PDAC), aiming to identify significant biomarkers that can serve as drug targets. We propose a systematic drug repurposing strategy that incorporates a deep neural network (DNN)-based drug-target interaction (DTI) model along with drug design specifications to develop a potential multi-molecule drug for PDAC treatment. We first established candidate protein-protein interaction networks and gene regulatory networks using big data mining techniques. Real PDAC and non-PDAC genome-wide genetic and epigenetic networks (GWGENs) were systematically identified using their corresponding microarray data through system identification and system order detection methods. The top 6,000 core GWGENs of PDAC and non-PDAC were extracted using the Principal Network Projection method. Subsequently, we annotated the core GWGENs using the Kyoto Encyclopedia of Genes and Genomes pathways to construct their respective core signaling pathways. By comparing upstream microenvironmental factors, core signaling pathways, and downstream aberrant cellular functions between PDAC and non-PDAC, we investigated the carcinogenic mechanisms of PDAC. Notably, c-MYC, forkhead box O3, and tumor suppressor p53 were identified as significant biomarkers for potential drug targets. Furthermore, the DNN-based DTI model predicted the interaction probabilities between candidate molecular drugs and these biomarkers. Based on drug design specifications such as regulatory ability, sensitivity, and toxicity, suitable multi-molecular potential drugs were selected. Ultimately, gemcitabine and MK-2206 were identified as a promising multi-molecular drug combination for PDAC treatment.

Keywords: Pancreatic cancer mechanisms; Systems biology; Big data mining; Genome-wide genetic and epigenetic networks; Kyoto Encyclopedia of Genes and Genomes pathways; Deep neural network-based drug-target interaction model; Drug design specifications; Principal network projection

***Corresponding author:**
Bor-Sen Chen
(bschen@ee.nthu.edu.tw)

Citation: Tsai Y, Chen B. Systemic drug repurposing for pancreatic cancer based on genetic and epigenetic network analysis using a systems biology approach and deep neural learning of drug-target interactions. *Tumor Discov*. 2025;4(1):47-67.
doi: 10.36922/td.4709

Received: August 30, 2024

Revised: October 8, 2024

Accepted: October 24, 2024

Published online: November 20, 2024

Copyright: © 2024 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC), often referred to as the "king of cancers," is the most common type of pancreatic cancer.[1] It is primarily characterized by a lack of significant symptoms in its early stages, making timely diagnosis challenging. Even when symptoms do emerge, they are frequently mistaken for other health conditions, typically indicating an advanced stage of the disease. Scholars predict that by 2030, pancreatic cancer will become the second leading cause of cancer-related deaths,[2] resulting in an estimated 46,000 deaths annually by 2040.[3] Despite numerous technological advancements in cancer treatment, the 5-year survival rate for PDAC patients remains only 12.2%.[4] A major contributing factor to this low survival rate is that pancreatic cancer is often diagnosed at an advanced stage, which severely limits the opportunities for surgical intervention. Compared to other cancers, pancreatic cancer has a relatively high mortality rate, influenced by various factors such as geographic region,[5] case numbers, and medical standards. The overall survival rate for PDCA is relatively low, primarily because tumors are often diagnosed at advanced stages and due to factors such as tumor location, size, and metastasis that affect treatment efficacy. Specific survival rate data may vary by region and treatment modality, reflecting the varying prevalence of pancreatic cancer across different populations. This variability is also influenced by various risk factors, including smoking, high-fat diets, obesity, and genetic factors.[6] Therefore, there is a pressing global need for enhanced preventive measures and genetic diagnostics to address this significant health challenge.[7]

The carcinogenic mechanisms underlying PDAC involve multiple gene mutations, abnormalities in signaling pathways, and the influence of the tumor microenvironment. Common gene mutations in PDAC include *KRAS*, *CDKN2A*, *TP53*, and *SMAD4*.[8] These mutations disrupt signaling pathways, leading to uncontrolled cell proliferation, inhibition of apoptosis, and promotion of tumor growth and metastasis.

Current treatment options for pancreatic cancer include surgical resection, radiation therapy, chemotherapy, immunotherapy, and targeted drugs. However, not all patients are suitable for surgery.[9] Common chemotherapy regimens often fail due to drug resistance, resulting from complex interactions among pancreatic cells, cancer cells, and the tumor microenvironment.[10] Consequently, clinical outcomes remain poor, and effective treatment methods are still lacking. Moreover, drug development typically takes at least 10 years and requires substantial funding.[11] Biopharmaceutical companies must invest heavily in research to identify drug targets, assess efficacy, consider side effects, and conduct extensive preclinical and clinical trials. Given the high failure rate in clinical trials,[12] there is a need for a more efficient drug development system. Recent advancements in deep learning have shown promise in drug discovery, with neural network methods being applied to predict drug-target interactions (DTI).[13-16] DTI data from various databases can help elucidate the relationships between drugs and their targets.[13,14] By framing drug and target features as a binary classification problem, DTI models based on deep neural networks (DNNs) can predict the interactions between drugs and targets (biomarkers), thereby identifying candidate multi-molecular drugs for specific diseases.[15,16] This approach, known as drug repurposing, utilizes existing drugs for new therapeutic purposes, potentially expediting the progression to preclinical and clinical trials.[17]

Before employing the DNN-based DTI model to predict potential multi-molecular drugs for PDAC biomarkers, we adopted a systems biology approach using whole-genome microarray data from PDAC and health controls to study their genome-wide genetic and epigenetic networks (GWGENs) for investigating the carcinogenic mechanisms of PDAC. The first step involved mining large databases to establish candidate GWGENs. Next, using system identification and systematic order detection techniques on the corresponding genome-wide microarray data of PDAC and healthy controls to eliminate false positive interactions and regulations, we obtained real GWGENs for PDAC and healthy controls. Given that the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations currently encompass only up to 6,000 nodes of GWGENs, the principal network projection (PNP) method was employed to individually extract the top 6,000 important nodes as core GWGENs from the real GWGENs of both PDAC and healthy controls. Then, we annotated the real GWGENs using KEGG pathways to construct core signaling pathways for both PDAC and healthy controls. By comparing the similarities and differences between the core signaling pathways of PDAC and healthy controls, along with their downstream cellular functional abnormalities, we investigated the carcinogenic mechanisms of PDAC and identified significant biomarkers, including c-MYC, forkhead box O3 (FOXO3), and tumor suppressor p53 (TP53), as potential drug targets. Finally, we combined the features of selected biomarkers with those of molecular drugs to train the DNN-based DTI model, predicting the probability of interaction between candidate molecular drugs and the identified drug targets (biomarkers). Based on drug design specifications, such as regulatory capacity, sensitivity, and toxicity, we identified potential drugs, ultimately selecting the combination of MK-2206 and gemcitabine as a promising multi-molecular drug approach to target key biomarkers for PDAC treatment.

2. Materials and methods

2.1. Overview of PDAC and healthy control genome-wide genetic and epigenetic networks using systems biology approach

In this study, we aim to establish the GWGENs of PDAC and non-PDAC core genomes. Microarray data for PDAC and non-PDAC were obtained from the National Center for Biotechnology Information (NCBI) under accession number GSE183795. Four processes were then conducted to identify the core signaling pathways of candidate GWGENs, as illustrated in Figure 1 and detailed below.

i. Construction of candidate GWGENs: We utilized a data mining approach to construct Boolean matrices representing candidate protein-protein interaction networks (PPINs) and candidate gene regulatory networks (GRNs), which include interactions among proteins, and regulation among genes, microRNAs (miRNAs), and long non-coding RNAs (lncRNAs). Specifically, if an interaction or regulation exists between two nodes, it is denoted as 1; if not, it is denoted as 0.
ii. Identification of real GWGENs: We employed PDAC and non-PDAC (control) microarray data to construct real GWGENs, identifying parameters for protein-protein interaction (PPI) models and GRN regulatory models by solving constrained linear least squares parameter estimation problems. To address potential false positive interactions in candidate GWGEN, we pruned these false positives using the Akaike Information Criterion (AIC) system order identification method, obtaining real GWGENs for PDAC and non-PDAC.
iii. Extraction of core GWGENs: We applied the PNP method to extract core GWGENs from the real GWGENs. The PNP method calculates the projection value of each node in the real GWGEN to capture 85% of the network's energy, sorting the projection values of all nodes from highest to lowest. Given the maximum allowable annotated node count of 6,000 (as per KEGG pathways), we selected the top 6,000 significant nodes to form the core GWGEN.
iv. Construction and comparison of core signaling pathways: We annotated the KEGG pathways of PDAC and non-PDAC of core GWGENs based on relevant literature, completing the construction of core signaling pathways for each. We then compared the upstream microenvironmental factors, core signaling pathways, and corresponding downstream aberrant cellular functions between PDAC and non-PDAC to explore the oncogenic molecular mechanisms of PDAC and identify potential biomarkers as drug targets for PDAC therapeutics.

2.2. Constructing candidate genome-wide genetic and epigenetic networks for PDAC and healthy controls based on big data mining

In this research, we obtained microarray data from the NCBI under accession number GSE183795. The dataset was divided into two groups: the disease group, comprising 139 PDAC samples, and the healthy control group, consisting of 105 non-PDAC samples.

The candidate GWGENs include candidate PPINs and candidate GRNs. We represented the candidate GWGEN using a binary Boolean matrix, where a value of 1 is assigned if an interaction or regulation exists for a node, and 0 if it does not. To construct the candidate PPINs, we consulted various databases, including the Database of Interacting Proteins (DIP),[18] IntAct,[19] the Biological General Repository for Interaction Datasets (BioGRID),[20] and the Molecular INTeraction Database (MINT).[21] For the candidate GRNs, we utilized multiple resources such as the Human Transcriptional Regulation Interaction Database (HTRIdb),[22] the integrated transcription factor platform (ITFP),[23] TRANSFAC,[24] CircuitDB,[25] TargetScanHuman,[26] and StarBase.[27]

2.3. Establishing a system model for identifying real genome-wide genetic and epigenetic networks for PDAC and healthy controls based on candidate genome-wide genetic and epigenetic networks

To investigate the oncogenic molecular mechanisms of PDAC, we referenced relevant databases and utilized PDAC microarray data to construct candidate GWGENs.[28] Following the establishment of these candidate GWGENs, we employed PDAC microarray data to discern the real GWGENs for PDAC and non-PDAC samples. This process required the development of a stochastic system model to enable candidate GWGENs to capture stochastic interactions and regulations, such as protein-protein interactions, as well as the regulation of transcription factors (TFs), miRNAs, and lncRNA. Additionally, the stochastic model should account for residuals from the initial model establishment and stochastic noise resulting from experimental measurements. Furthermore, the main protein interaction model in Equation I and the miRNA regulation models in Equations II–IV were designed as bilinear interaction models based on the product of the concentrations of the interacting proteins in Equation I or the regulations of miRNAs on their target mRNAs, miRNAs or lncRNAs in Equations II–IV. However, for simplification, we presented the interaction and regulation coefficients as linear in PPINs and GRNs.[28]

First, we established a system model of the interactions involving the w-th protein and other proteins in the candidate PPINs, presented as follows:

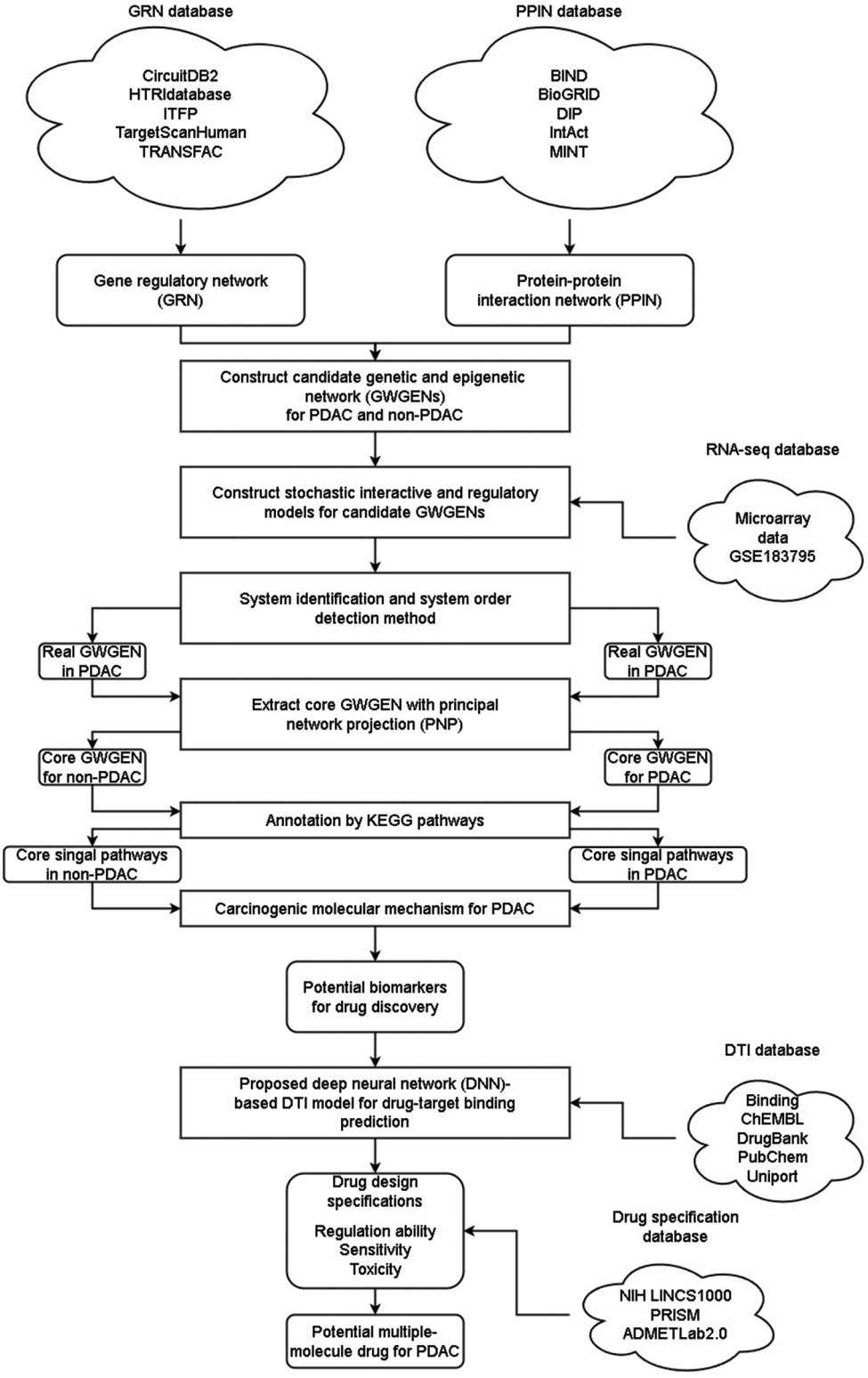


Figure 1. Flowchart of the systems biology approach to investigate the carcinogenic molecular mechanisms and systematic drug design for PDAC
Abbreviations: BioGRID: Biological General Repository for Interaction Datasets; DIP: Database of Interacting Proteins; DTI: Drug-target interaction; HTRI: Human Transcriptional Regulation Interaction; ITFP: Integrated Transcription Factor Platform; KEGG: Kyoto Encyclopedia of Genes and Genomes; MINT: Molecular INTeraction Database; PDAC: Pancreatic ductal adenocarcinoma; PRISM: Pharmaceutical Regulatory Information System.

$$p_w[n] = \sum_{r=1}^{R_w} \pi_{w-r} p_w[n] p_r[n] + \beta_{w-PPIN} + \delta_{w-PPIN}[n] \quad \text{(I)}$$

for w = 1,2…,W-1,W，n = 1,2…,N-1,N

among which p_w [n] and p_r [n] denote the expression levels of the w-th protein and the r-th protein in the n-th sample, respectively; π_{w_r} represents the interaction ability between the r-th protein and the w-th protein; R_w is the total number of proteins interacting with the w-th protein; W denotes the total number of proteins in the candidate PPIN; N is the number of data samples (PDAC or non-PDAC); β_{w_PPIN} indicates the basal expression level of the w-th protein, reflecting changes due to interactions such as acetylation, phosphorylation, or unknown histone modifications that cannot be modeled in (1); δ_{w_PPIN} [n] accounts for random noise affecting the w-th protein in the n-th sample due to modeling residuals, experimental measurement errors, or environmental interference.

After establishing the PPI system model, we also developed a regulatory system model to describe the relationships between genes and their regulators, including TFs, miRNAs, and lncRNAs. The transcriptional regulation system model for the x-th gene in the n-th sample can be expressed as:[28]

$$g_x[n] = \sum_{s=1}^{S_x} \alpha_{x-s} t_s[n] + \sum_{u=1}^{U_x} \gamma_{x-u} l_u[n] - \sum_{v=1}^{V_x} \varepsilon_{x-v} m_v[n] g_x[n] + \beta_x + \delta_x[n] \quad \text{(II)}$$

for x = 1,2…,X-1,X, n = 1,2…,N-1,N

where g_x [n] indicates the expression level of the x-th gene in the n-th sample; S_x, U_x, and V_x represent the total number of TFs, lncRNAs, and miRNAs binding to the x-th gene, respectively; t_s [n], l_n [n], and m_v [n] denote the expression levels of the s-th TF, the u-th lncRNA, and the v-th miRNA in the n-th sample, respectively; α_{x_s} symbolizes the transcriptional regulatory ability of the s-th TF on the x-th gene; γ_{x_u} signifies the transcriptional regulatory ability of the u-th lncRNA on the x-th gene; $\varepsilon_{x_y} > 0$ indicates the post-transcriptional regulatory ability of the v-th miRNA in degrading the x-th gene's miRNA; X is the total number of genes in the candidate GWGEN; N is the number of data samples (PDAC or non-PDAC); β_x represents the basal expression level of the x-th gene, influenced by modifications such as methylation, phosphorylation, acetylation, or unknown gene regulatory effects; δ_x [n] accounts for the random noise affecting the x-th gene in the n-th sample due to modeling residuals and measurement errors.

Next, we established the regulatory system model for lncRNA. The transcriptional regulation system model for the y-th lncRNA in the n-th sample is described as:

$$l_y[n] = \sum_{s=1}^{S_y} \zeta_{y-s} t_s[n] + \sum_{u=1}^{U_y} \psi_{y-u} l_u[n] - \sum_{v=1}^{V_y} \kappa_{y-v} m_v[n] l_y[n] + \beta_y + \delta_y[n] \quad \text{(III)}$$

for y = 1,2…,Y-1,Y, n = 1,2…,N-1,N

where l_y [n] indicates the expression level of the y-th lncRNA in the n-th sample; S_y, U_y, and V_y represent the total numbers of TFs, lncRNAs, and miRNAs binding to the y-th lncRNA, respectively; t_s [n], l_u [n], and m_v [n] denote the expression levels of the s-th TF, the u-th lncRNA, and the v-th miRNA in the n-th sample, respectively; ζ_{y_s} symbolizes the transcriptional regulatory ability of the s-th TF on the y-th lncRNA; ψ_{y_u} signifies the transcriptional regulatory ability of the u-th lncRNA on the y-th lncRNA; $k_{y_v} > 0$ represents the post-transcriptional regulatory ability of the v-th miRNA in degrading the y-th lncRNA's miRNA; Y is the total number of lncRNAs in the candidate GWGEN; N is the number of data samples (PDAC or non-PDAC); β_y reflects the expression basal level of the y-th lncRNA, influenced by modifications such as phosphorylation, acetylation, or unknown gene regulatory effects; δ_y [n] captures the random noise affecting the y-th lncRNA in the n-th sample due to modeling residuals and measurement errors.

Finally, we established the miRNA regulatory system model in a similar manner. The transcriptional regulation system model for the z-th miRNA in the n-th sample can be described by the following equation:

$$m_z[n] = \sum_{s=1}^{S_z} \lambda_{z-s} t_s[n] + \sum_{u=1}^{U_z} \mu_{z-u} l_u[n] - \sum_{v=1}^{V_z} \rho_{z-v} m_v[n] m_z[n] + \beta_z + \delta_z[n] \quad \text{(IV)}$$

for z = 1,2…,Z-1,Z, n = 1,2…,N-1,N

where m_z [n] represents the expression level of the z-th miRNA in the n-th sample; S_z, U_z, and V_z indicate the total number of TFs, lncRNAs, and miRNAs binding to the z-th miRNA, respectively; t_s [n], l_u [n], and m_v [n] denote the expression levels of the s-th TF, the u-th lncRNA, and the v-th miRNA in the n-th sample, respectively; λ_{z_s} symbolizes the transcriptional regulatory ability of the s-th TF on the z-th miRNA; μ_{z_u} signifies the transcriptional regulatory ability of the u-th lncRNA on the z-th miRNA; $\rho_{z_v} > 0$ represents the post-transcriptional regulatory ability of the v-th miRNA in degrading the z-th miRNA's miRNA; Z is the total number of miRNAs in the candidate GWGEN; N is the number of data samples (PDAC or non-PDAC); β_z indicates the change in the expression of the z-th miRNA due to phosphorylation, acetylation, or unknown gene regulatory effects; δ_z [n] accounts for the random noise affecting the z-th miRNA in the n-th sample due to model residuals and experimental measurement errors.

2.4. Elimination of false positives from candidate genome-wide genetic and epigenetic networks for real genome-wide genetic and epigenetic networks of PDAC and healthy controls based on system identification and order detection methods

Based on the stochastic system models of protein interaction and gene regulations described above, we constructed four interaction and regulation models for candidate GWGENs, including the candidate PPI model, candidate gene regulation model, candidate lncRNA regulation model, and candidate miRNA regulation model in Equations I–IV. Next, we pruned false positive interactions and regulations, from these candidate GWGENs using system identification and order detection methods, utilizing microarray data from both PDAC and non-PDAC samples to estimate the corresponding real GWGENs.

Equations I–IV can be reformulated as the following linear regression equations to obtain parameter vectors for interactions and gene regulations.

$$p_w[n]=\left[p_w[n]p_1[n]p_w[n]p_2[n]\cdots p_w[n]p_{R_w}[n]1\right]\times\begin{bmatrix}\pi_{w-1}\\ \pi_{w-2}\\ \vdots\\ \pi_{w-R_w}\\ \beta_{w-PPIN}\end{bmatrix}+\delta_{w-PPIN}[n]=\boldsymbol{\phi}_w[n]\cdot\theta_w+\delta_{w-PPIN}[n] \quad (V)$$

for w =1,2…,W-1,W, n = 1,2…,N-1,N

$$g_x[n]=\begin{bmatrix}t_1[n]\cdots t_{S_x}[n]\,l_1[n]\\ \cdots l_{U_x}[n]\,m_1[n]g_x[n]\cdots\\ m_{V_x}[n]g_x[n]1\end{bmatrix}\times\begin{bmatrix}\alpha_{x-1}\\ \vdots\\ \alpha_{x-S_x}\\ \gamma_{x-1}\\ \vdots\\ \gamma_{x-U_x}\\ -\varepsilon_{x-1}\\ \vdots\\ -\varepsilon_{x-V_x}\\ \beta_x\end{bmatrix}+\delta_x[n]=\phi_x[n]\cdot\theta_x+\delta_x[n] \quad (VI)$$

for x = 1,2…,X-1,X, n = 1,2…,N-1,N

$$l_y[n]=\begin{bmatrix}t_1[n]\cdots t_{S_y}[n]\,l_1[n]\cdots l_{U_y}\\ [n]\,m_1[n]\,l_y[n]\cdots\\ m_{V_y}[n]l_y[n]\,1\end{bmatrix}\times\begin{bmatrix}\zeta_{y-1}\\ \vdots\\ \zeta_{y-S_y}\\ \psi_{y-1}\\ \vdots\\ \psi_{y-U_y}\\ -\kappa_{y-1}\\ \vdots\\ -\kappa_{y-V_y}\\ \beta_y\end{bmatrix}+\delta_y[n]=\phi_y[n]\cdot\theta_y+\delta_y[n] \quad (VII)$$

for y =1,2…,Y-1,Y, n = 1,2…,N-1,N

$$m_z[n]=\begin{bmatrix}t_1[n]\cdots t_{S_z}[n]\,l_1[n]\cdots l_{U_z}\\ [n]\,m_1[n]\,m_z[n]\cdots\\ m_{V_z}[n]\,m_z[n]\,1\end{bmatrix}\times\begin{bmatrix}\lambda_{z-1}\\ \vdots\\ \lambda_{z-S_z}\\ \mu_{z-1}\\ \vdots\\ \mu_{z-U_z}\\ -\rho_{z-1}\\ \vdots\\ -\rho_{z-V_z}\\ \beta_z\end{bmatrix}+\delta_z[n]=\phi_z[n]\cdot\theta_z+\delta_z[n] \quad (VIII)$$

for z =1,2…,Z-1,Z, n = 1,2…,N-1,N

Next, with N representing microarray data samples, we can express these as linear Equations V–VIII:

$$\begin{bmatrix} p_w[1] \\ p_w[2] \\ \vdots \\ p_w[N] \end{bmatrix} = \begin{bmatrix} \phi_w[1] \\ \phi_w[2] \\ \vdots \\ \phi_w[N] \end{bmatrix} \theta_w + \begin{bmatrix} \delta_w[1] \\ \delta_w[2] \\ \vdots \\ \delta_w[N] \end{bmatrix} \quad \text{(IX)}$$

for w =1,2…,W-1,W, n = 1,2…,N-1,N

$$\begin{bmatrix} g_x[1] \\ g_x[2] \\ \vdots \\ g_x[N] \end{bmatrix} = \begin{bmatrix} \phi_x[1] \\ \phi_x[2] \\ \vdots \\ \phi_x[N] \end{bmatrix} \theta_x + \begin{bmatrix} \delta_x[1] \\ \delta_x[2] \\ \vdots \\ \delta_x[N] \end{bmatrix} \quad \text{X)}$$

for x = 1,2…,X-1,X, n = 1,2…,N-1,N

$$\begin{bmatrix} l_y[1] \\ l_y[2] \\ \vdots \\ l_y[N] \end{bmatrix} = \begin{bmatrix} \phi_y[1] \\ \phi_y[2] \\ \vdots \\ \phi_y[N] \end{bmatrix} \theta_y + \begin{bmatrix} \delta_y[1] \\ \delta_y[2] \\ \vdots \\ \delta_y[N] \end{bmatrix} \quad \text{(XI)}$$

for y =1,2…,Y-1,Y, n = 1,2…,N-1,N

$$\begin{bmatrix} m_z[1] \\ m_z[2] \\ \vdots \\ m_z[N] \end{bmatrix} = \begin{bmatrix} \phi_z[1] \\ \phi_z[2] \\ \vdots \\ \phi_z[N] \end{bmatrix} \theta_z + \begin{bmatrix} \delta_z[1] \\ \delta_z[2] \\ \vdots \\ \delta_z[N] \end{bmatrix} \quad \text{(XII)}$$

for z = 1,2…,Z-1,Z, n = 1,2…,N-1,N

Further Equations IX–XII can be represented by the following algebraic equations individually:

$$P_w = \Phi_w \cdot \Theta_w + \Delta_w \ \ for\ w = 1,2\ldots,W-1,W \quad \text{(XIII)}$$

$$G_x = \Phi_x \cdot \Theta_x + \Delta_x \ \ for\ x = 1,2\ldots,X-1,X \quad \text{(XIV)}$$

$$L_y = \Phi_y \cdot \Theta_y + \Delta_y \ \ for\ y = 1,2\ldots,Y-1,Y \quad \text{(XV)}$$

$$M_z = \Phi_z \cdot \Theta_z + \Delta_z \ \ for\ z = 1,2\ldots,Z-1,Z \quad \text{(XVI)}$$

where Φ_w is the linear regression matrix for proteins, Φ_x is the linear regression matrix for genes, Φ_y is the linear regression matrix for lncRNAs, and Φ_z is the linear regression matrix for miRNAs.

Next, we employed a constrained linear least-squares parameter estimation method to estimate parameter vectors Φ_w, Φ_x, Φ_y and Φ_z. Specifically, we imposed constraints to ensure that the degrading effects of miRNAs on post-transcriptional genes, lncRNAs, and miRNAs are negative. The parameter estimation problem for the GWGENs of PDAC and non-PDAC can be solved by the following constrained least-squares parameter estimation problem equations:

$$\hat{\Theta}_w = \underset{\Theta_w}{\operatorname{argmin}} \frac{1}{2} \Phi_w \cdot \Theta_w - P_{w2}^2 \quad \text{(XVII)}$$

$$\hat{\Theta}_x = \underset{\Theta_x}{\operatorname{argmin}} \frac{1}{2} \Phi_x \cdot \Theta_x - G_{x2}^2 \quad \text{(XVIII)}$$

$$subject\ to \left[\begin{array}{ccc|ccc|ccc|c} 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{S_x}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{U_x}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{V_x}} & \end{array}\right] \cdot \Theta_x \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\hat{\Theta}_y = \underset{\Theta_y}{\operatorname{argmin}} \frac{1}{2} \Phi_y \cdot \Theta_y - L_{y2}^2 \quad \text{(XIX)}$$

$$subject\ to \left[\begin{array}{ccc|ccc|ccc|c} 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{S_y}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{U_y}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{V_y}} & \end{array}\right] \cdot \Theta_y \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\hat{\Theta}_z = \underset{\Theta_z}{\operatorname{argmin}} \frac{1}{2} \Phi_z \cdot \Theta_z - M_{z2}^2 \quad \text{(XX)}$$

$$subject\ to \left[\begin{array}{ccc|ccc|ccc|c} 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \\ \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{S_z}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{U_z}} & \multicolumn{3}{c|}{\underbrace{\qquad\qquad}_{V_z}} & \end{array}\right] \cdot \Theta_z \le \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Given that the regulatory effects of miRNAs on post-transcriptional genes, lncRNAs, and other miRNAs must be negative, we utilized the MATLAB Optimization Toolbox to solve the constrained least-squares parameter estimation problems with their added constraints in Equations XVII–XX. This approach allowed us to derive optimal estimated parameter vectors for PPIs, as well as for gene, lncRNA, and miRNA regulations within the GWGENs for both PDAC and non-PDAC.

To address the issue of numerous false positive interactions identified among the candidate GWGENs, we

employed the systematic order detection method using the AIC to prune these inaccuracies and derive real GWGENs for both PDAC and non-PDAC.[29]

The system order detection scheme, based on the AIC method, estimates the number of interactions among proteins and the number of regulations involving genes, lncRNAs, and miRNAs within the candidate GWGENs as follows:

$$AIC\left(R_w\right)=\log\left(\Omega_w^2\right)+\frac{2\left(1+R_w\right)}{N} \quad \text{(XXI)}$$

$$\text{where } \Omega_w^2=\sqrt[2]{\frac{\left(\Phi_w\cdot\hat{\Theta}_w-P_w\right)^T\left(\Phi_w\cdot\hat{\Theta}_w-P_w\right)}{N}}$$

where the parameter vector $\hat{\Theta}_w$ can be obtained from Equation (XVII) using the least-squares method. Ω_w represents the residual estimation of the w-th protein model, and R_w is the number of interactions (i.e., the model's complexity) with the w-th protein.

$$AIC\left(S_x,U_x,V_x\right)=\log\left(\Omega_x^2\right)+\frac{2\left(S_x+U_x+V_x+1\right)}{N} \quad \text{(XXII)}$$

$$\textit{where } \Omega_x^2=\sqrt[2]{\frac{\left(\Phi_x\cdot\hat{\Theta}_x-G_x\right)^T\left(\Phi_x\cdot\hat{\Theta}_x-G_x\right)}{N}}$$

where the parameter vector $\hat{\Theta}_x$ can be obtained from Equation (XVIII) using the least-squares method. Ω_x represents the residual estimation of the x-th gene model, while S_x, U_x, and V_x represent the number of gene, lncRNA, and miRNA regulations acting on the x-th gene, respectively.

$$AIC\left(S_y,U_y,V_y\right)=\log\left(\Omega_y^2\right)+\frac{2\left(S_y+U_y+V_y+1\right)}{N} \quad \text{(XXIII)}$$

$$\textit{with } \Omega_y^2=\sqrt[2]{\frac{\left(\Phi_y\cdot\hat{\Theta}_y-L_y\right)^T\left(\Phi_y\cdot\hat{\Theta}_y-L_y\right)}{N}}$$

where the parameter vector $\hat{\Theta}_y$ can be obtained from Equation (XIX) using the least-squares method. Ω_y represents the residual estimation of the y-th lncRNA model, whereas S_y, U_y, and V_y represent the number of genes, lncRNA, and miRNA regulations acting on the y-th lncRNA, respectively.

$$AIC\left(S_z,U_z,V_z\right)=\log\left(\Omega_z^2\right)+\frac{2\left(S_z+U_z+V_z+1\right)}{N} \quad \text{(XXIV)}$$

$$\textit{where } \Omega_z^2=\sqrt[2]{\frac{\left(\Phi_z\cdot\hat{\Theta}_z-M_z\right)^T\left(\Phi_z\cdot\hat{\Theta}_z-M_z\right)}{N}}$$

where the parameter vector $\hat{\Theta}_z$ can be obtained from Equation XX using the least-squares method. Ω_z represents the residual estimation of the z-th miRNA model, and S_z, U_z, and V_z respectively represent the number of gene, lncRNA, and miRNA regulations acting on the z-th miRNA, respectively.

The AIC method was used to assess the complexity of stochastic models and their fit to data. A smaller value of AIC indicates a better model fit while accounting for the model's complexity. To obtain the real GWGEN, we had to minimize four AIC Equations XXI–XXIV:

$$R_w^*=\underset{R_w}{\operatorname{argmin}}\,AIC\left(R_w\right) \quad \text{(XXV)}$$

$$S_x^*,U_x^*,V_x^*=\underset{S_x,U_x,V_x}{\operatorname{argmin}}\,AIC\left(S_x,U_x,V_x\right) \quad \text{(XXVI)}$$

$$S_y^*,U_y^*,V_y^*=\underset{S_y,U_y,V_y}{\operatorname{argmin}}\,AIC\left(S_y,U_y,V_y\right) \quad \text{(XXVII)}$$

$$S_z^*,U_z^*,V_z^*=\underset{S_z,U_z,V_z}{\operatorname{argmin}}\,AIC\left(S_z,U_z,V_z\right) \quad \text{(XXVIII)}$$

R_w^* represents the actual quantity of protein interactions with the w-th protein. S_x^*, U_x^*, and V_x^* represent the actual number of regulatory TFs, lncRNAs, and miRNAs on the x-th gene, respectively. S_y^*, U_y^*, and V_y^* indicate the actual quantities of regulations between TFs, lncRNAs, and miRNAs on the y-th lncRNA, respectively. S_z^*, U_z^*, and V_z^* denote the actual number of regulatory TFs, lncRNAs, and miRNAs on the z-th miRNA, respectively. By minimizing the AIC, we successfully identified the true number of interactions for each protein and the actual number of regulations for each gene, miRNA and lncRNA, effectively pruning false positives from the candidate GWGENs, and thus obtaining the real GWGENs for both PDAC and non-PDAC, as depicted in Figure S1.

2.5. Extracting core genome-wide genetic and epigenetic networks from real genome-wide genetic and epigenetic networks using the principal network projection method

By employing the AIC method to prune false positive interactions and regulations from candidate GWGENs, real GWGENs for both PDAC and non-PDAC were obtained. However, due to the complexity of real GWGENs from both PDAC and non-PDAC, the carcinogenic molecular mechanisms of PDAC have not yet been fully confirmed. Therefore, it was necessary to leverage KEGG pathways to elucidate the signaling pathways underlying real GWGENs of both PDAC and non-PDAC. Since the current KEGG database can only annotate networks containing up to 6,000 nodes,

we focused on extracting the top 6,000 significant nodes (i.e., core GWGENs) from both the PDAC and non-PDAC real GWGENs using the PNP method. Subsequently, KEGG pathways were utilized to annotate these core GWGENs.

The PNP method we employed involves performing singular value decomposition (SVD) on the real GWGENs, as shown in Figure S1, followed by extracting the top 6,000 ranked nodes to construct the core GWGENs, as depicted in Figure S2. To perform SVD, we first constructed the network matrix *K* for the real GWGEN:

$$K=\begin{bmatrix} k_{protein\leftrightarrow protein} & 0 & 0 \\ k_{TF\rightarrow gene} & k_{lncRNA\rightarrow gene} & k_{miRNA\rightarrow gene} \\ k_{TF\rightarrow lncRNA} & k_{lncRNA\rightarrow lncRNA} & k_{miRNA\rightarrow lncRNA} \\ k_{TF\rightarrow miRNA} & k_{lncRNA\rightarrow miRNA} & k_{miRNA\rightarrow miRNA} \end{bmatrix} \quad \text{(XXIX)}$$

$$K\in\mathbb{R}^{(W+X+Y+Z)\times(X+Y+Z)}$$

where the submatrix $k_{protein\leftrightarrow protein}$ represents the estimated interaction abilities between proteins in the PPIN. Since protein interactions are bidirectional, these are represented by double-headed arrows. The respective submatrices $k_{TF\rightarrow gene}$, $k_{lncRNA\rightarrow gene}$, and $k_{miRNA\rightarrow gene}$ represent the estimated regulatory network of TFs, lncRNAs, and miRNAs that regulate or transcribe genes. Additionally, the submatrices $k_{TF\rightarrow lncRNA}$, $k_{lncRNA\rightarrow lncRNA}$, and $k_{miRNA\rightarrow lncRNA}$ represent estimated networks of TFs, lncRNAs, and miRNAs involved in regulating or transcribing lncRNAs, respectively. Finally, the submatrices $k_{TF\rightarrow miRNA}$, $k_{lncRNA\rightarrow miRNA}$, and $k_{miRNA\rightarrow miRNA}$ represent the estimated regulations of TFs, lncRNAs, and miRNAs in miRNA transcription, respectively.

Below is the detailed description of the network matrix *K* for real GWGENs of PDAC and non-PDAC:

$$K=\begin{bmatrix} \hat{\pi}_{1-1} & \cdots & \hat{\pi}_{1-r} & \cdots & \hat{\pi}_{1-Rw} & & 0 & \cdots & 0 & \cdots & 0 & & 0 & \cdots & 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \iddots & \vdots & & \vdots & \ddots & \vdots & \iddots & \vdots & & \vdots & \ddots & \vdots & \iddots & \vdots & \\ \hat{\pi}_{w-1} & \cdots & \hat{\pi}_{w-r} & \cdots & \hat{\pi}_{w-Rw} & & 0 & \cdots & 0 & \cdots & 0 & & 0 & \cdots & 0 & \cdots & 0 & \\ \vdots & \iddots & \vdots & \ddots & \vdots & & \vdots & \iddots & \vdots & \ddots & \vdots & & \vdots & \iddots & \vdots & \ddots & \vdots & \\ \hat{\pi}_{W-1} & \cdots & \hat{\pi}_{W-r} & \cdots & \hat{\pi}_{W-Rw} & & 0 & \cdots & 0 & \cdots & 0 & & 0 & \cdots & 0 & \cdots & 0 & \\ \hat{\alpha}_{1-1} & \cdots & \hat{\alpha}_{1-s} & \cdots & \hat{\alpha}_{1-Sx} & \hat{\gamma}_{1-1} & \cdots & \hat{\gamma}_{1-u} & \cdots & \hat{\gamma}_{1-Ux} & -\hat{\varepsilon}_{1-1} & \cdots & -\hat{\varepsilon}_{1-v} & \cdots & -\hat{\varepsilon}_{1-Vx} \\ \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ \hat{\alpha}_{x-1} & \cdots & \hat{\alpha}_{x-s} & \cdots & \hat{\alpha}_{x-Sx} & \hat{\gamma}_{x-1} & \cdots & \hat{\gamma}_{x-u} & \cdots & \hat{\gamma}_{x-Ux} & -\hat{\varepsilon}_{x-1} & \cdots & -\hat{\varepsilon}_{x-v} & \cdots & -\hat{\varepsilon}_{x-Vx} \\ \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ \hat{\alpha}_{X-1} & \cdots & \hat{\alpha}_{X-s} & \cdots & \hat{\alpha}_{X-Sx} & \hat{\gamma}_{X-1} & \cdots & \hat{\gamma}_{X-u} & \cdots & \hat{\gamma}_{X-Ux} & -\hat{\varepsilon}_{X-1} & \cdots & -\hat{\varepsilon}_{X-v} & \cdots & -\hat{\varepsilon}_{X-Vx} \\ \hat{\zeta}_{1-1} & \cdots & \hat{\zeta}_{1-s} & \cdots & \hat{\zeta}_{1-Sy} & \hat{\psi}_{1-1} & \cdots & \hat{\psi}_{1-u} & \cdots & \hat{\psi}_{1-Uy} & -\hat{\kappa}_{1-1} & \cdots & -\hat{\kappa}_{1-v} & \cdots & -\hat{\kappa}_{1-Vy} \\ \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ \hat{\zeta}_{y-1} & \cdots & \hat{\zeta}_{y-s} & \cdots & \hat{\zeta}_{y-Sy} & \hat{\psi}_{y-1} & \cdots & \hat{\psi}_{y-u} & \cdots & \hat{\psi}_{y-Uy} & -\hat{\kappa}_{y-1} & \cdots & -\hat{\kappa}_{y-v} & \cdots & -\hat{\kappa}_{y-Vy} \\ \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ \hat{\zeta}_{Y-1} & \cdots & \hat{\zeta}_{Y-s} & \cdots & \hat{\zeta}_{Y-Sy} & \hat{\psi}_{Y-1} & \cdots & \hat{\psi}_{Y-u} & \cdots & \hat{\psi}_{Y-Uy} & -\hat{\kappa}_{Y-1} & \cdots & -\hat{\kappa}_{Y-v} & \cdots & -\hat{\kappa}_{Y-Vy} \\ \hat{\lambda}_{1-1} & \cdots & \hat{\lambda}_{1-s} & \cdots & \hat{\lambda}_{1-Sz} & \hat{\mu}_{1-1} & \cdots & \hat{\mu}_{1-u} & \cdots & \hat{\mu}_{1-Uz} & -\hat{\rho}_{1-1} & \cdots & -\hat{\rho}_{1-v} & \cdots & -\hat{\rho}_{1-Vz} \\ \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \iddots & \vdots \\ \hat{\lambda}_{z-1} & \cdots & \hat{\lambda}_{z-s} & \cdots & \hat{\lambda}_{z-Sz} & \hat{\mu}_{z-1} & \cdots & \hat{\mu}_{z-u} & \cdots & \hat{\mu}_{z-Uz} & -\hat{\rho}_{z-1} & \cdots & -\hat{\rho}_{z-v} & \cdots & -\hat{\rho}_{z-Vz} \\ \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \ddots & \vdots \\ \hat{\lambda}_{Z-1} & \cdots & \hat{\lambda}_{Z-s} & \cdots & \hat{\lambda}_{Z-Sz} & \hat{\mu}_{Z-1} & \cdots & \hat{\mu}_{Z-u} & \cdots & \hat{\mu}_{Z-Uz} & -\hat{\rho}_{Z-1} & \cdots & -\hat{\rho}_{Z-v} & \cdots & -\hat{\rho}_{Z-Vz} \end{bmatrix} \quad \text{(XXX)}$$

Next, we applied SVD to the network matrix *K* for the real GWGENs of PDAC and non-PDAC as follows:

$$\boldsymbol{K}=\boldsymbol{\mathcal{U}}\sum\boldsymbol{\mathcal{V}}^{\boldsymbol{T}}$$

$$\boldsymbol{\mathcal{U}}\in\mathbb{R}^{(\boldsymbol{W}+\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})\times(\boldsymbol{W}+\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})}$$

$$\sum\in\mathbb{R}^{(\boldsymbol{W}+\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})\times(\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})}$$

$$\boldsymbol{\mathcal{V}}\in\mathbb{R}^{(\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})\times(\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z})} \quad \text{(XXXI)}$$

$$where\ \sum=\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{(X+Y+Z)} \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \quad \text{(XXXII)}$$

in which the singular values σ_I are arranged in decreasing order, i.e., $\sigma^1 \geq \sigma_2 \ldots \geq \sigma_{(X+Y+Z)} \geq 0$.

Based on energy considerations, we chose to retain the first *J* singular values from the singular matrix Σ in network matrix *K*, ensuring they account for at least 85% of the total energy in the real GWGEN.

Subsequently, we retained the first *J* rows of the singular matrices *u* and *v*, establishing the significant structural component of the network, which contains at least 85% of the total energy in the real GWGEN,[30] as indicated below:

$$E_j=\frac{\sum_{i=1}^{J}\sigma_i^2}{\sum_{c=1}^{X+Y+Z}\sigma_c^2}\geq 0.85 \quad \text{(XXXIII)}$$

Next, we projected each row of the network matrix *K* (representing the interactions or regulations of each node in the real GWGEN) onto the first *J* significant singular vectors. The 2-norm of the projection value for each protein, gene, miRNA, and lncRNA node corresponds to the first *J* principal singular vectors of both PDAC and non-PDAC real GWGENs as described below:

$$Project\left(a,b\right)=K_a\cdot V_b^T$$

$$P_{2-norm}\left(a\right)=\sqrt[2]{\sum_{b=1}^{J}Project\left(a,b\right)^2} \quad \text{(XXXIV)}$$

for a = 1,2…, (W + X + Y + Z), b =1,2…, J-1, J

where *Project*(*a*,*b*) represents the projection value of the *a*-th node onto the *b*-th principal singular vector;

K_a denotes the a-th row of the matrix K; V_b indicates the b-th principal singular vector; P_{2_norm} (a) represents the square root of the sum of the squared projection values of the a-th node onto the first J singular vectors, reflecting the significance of the node within the real GWGEN from the network energy perspective.

Using the 2-norm projection value P_{2_norm} (a), we extracted the top 6,000 ranked nodes, which we classified as important nodes to construct the core GWGENs for PDAC and non-PDAC, as shown in Figure S2. We then utilized KEGG pathways to annotate the core GWGENs of both PDAC and non-PDAC, helping to identify the core signaling pathways for both conditions, as presented in Figure 2. In addition, considering information loss of real GWGEN, we actually obtained nearly identical signaling pathways of PDAC under different SVD truncation thresholds (75%, 85%, and 95%), demonstrating that filtering off insignificant singular values did not substantially affect the KEGG pathway annotations. By comparing the core signaling pathways of PDAC and non-PDAC in Figure 2, we aim to investigate the carcinogenic mechanisms of PDAC. Based on these insights, we identified key biomarkers, listed in Table 1, which may serve as potential drug targets for drug repurposing to treat PDAC.

2.6. Predicting drug candidates using deep neural network-based drug-target interaction model and screening by design specifications for treating PDAC

After identifying three important biomarkers implicated in the carcinogenic mechanisms of PDAC as candidate drug targets, we trained a DNN-DTI model to predict potential molecular drugs targeting these biomarkers. We utilized drug-target interaction data from several databases such as KEGG,[31] BIDD,[32] UniProt,[33] DrugBank,[34] PubChem,[35] ChEMBL,[36] and STITCH.[37] After identifying potential molecular drugs for PDAC and considering their regulatory ability, sensitivity, and toxicity as design specifications, we proceeded with drug repurposing and design.

Before training the DNN-DTI model, we preprocessed the DTI data. We gathered relevant interaction data from the aforementioned databases and converted the drug-

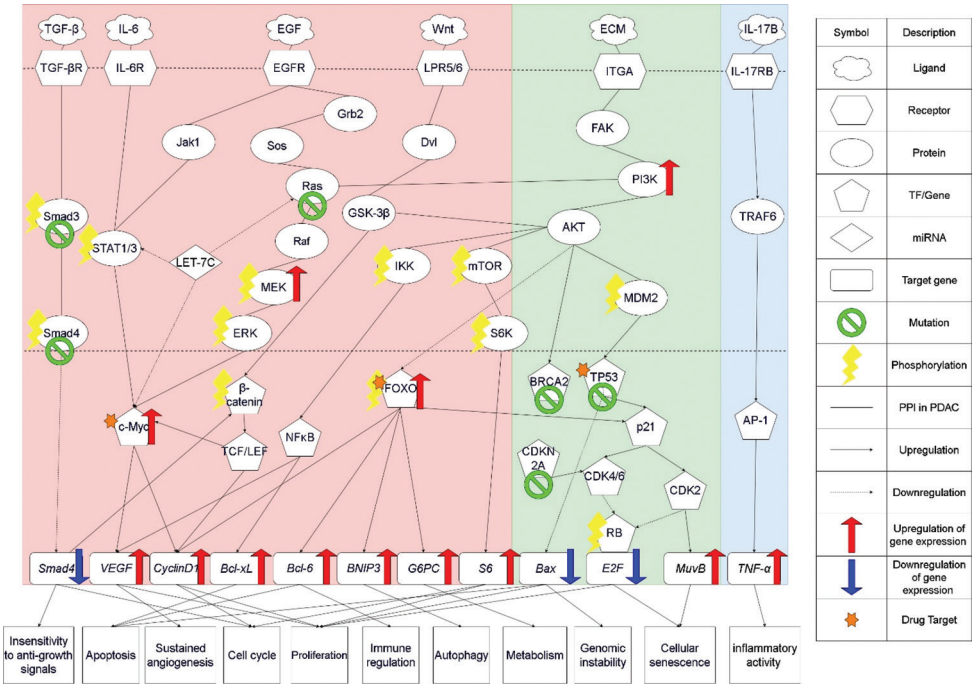


Figure 2. Shared and distinct core signaling pathways and downstream cellular functional impairments between PDAC and healthy controls
Abbreviations: AKT: Protein kinase B; AP-1: Activator protein 1; BNIP: BCL2 interacting protein 3; BRCA: Breast cancer gene; CDK: Cyclin-dependent kinase; ECM: Extracellular matrix; EGF: Epidermal growth factor; EGFR: Epidermal growth factor receptor; ERK: Extracellular signal-regulated kinase 1; FAK: Focal adhesion kinase; FOXO: Forkhead box O; GSK: Glycogen synthase kinase; G6PC: Glucose-6-phosphatase; IKK: IκB kinase; IL: Interleukin; IL-XR: Interleukin X receptor; ITGA: Integrin alpha; Jak1: Janus kinase 1; LEF: Lymphoid enhancer factor; LPR: Low-density-lipoprotein-receptor-related-protein; MDM2: Mouse double minute 2; MEK: Mitogen-activated extracellular signal-regulated kinase; mTOR: Mammalian target of rapamycin; NFκB: Nuclear factor κ B; PDAC: Pancreatic ductal adenocarcinoma; PI3K: Phosphoinositide 3-kinase; PPI: Protein-protein interaction; RB: Retinoblastoma protein; STAT: Signal transducer and activator of transcription; S6K: Ribosomal protein S6 kinase; TCF: T cell factor; TF: Transcription factor; TGF-β: Transforming growth factor β; TβR: Transforming growth factor β receptor; TNF-α: Tumor necrosis factor α; TRAF6: Tumor necrosis factor receptor-associated factor 6; VEGF: Vascular endothelial growth factor.

Table 1. Information on candidate molecular drugs for selected pancreatic ductal adenocarcinoma biomarkers based on their regulatory capacity, sensitivity, and toxicity

Target biomarker : c-MYC(+)			
Potential drug	**Regulation ability (L1000)**	**Sensitivity (PRISM)**	**Toxicity (LC50, mol/kg)**
Tipranavir	−1.33845971	−0.283941421	4.556
Tolcapone	−0.072789862	−0.26251132	4.78
Gemcitabine	−0.544798394	−2.417963872	2.381
Target biomarker : FOXO3(+)			
Potential drug	**Regulation ability (L1000)**	**Sensitivity (PRISM)**	**Toxicity (LC50, mol/kg)**
Atracurium	−0.238786879	−0.468347976	5.587
MK-2206	−0.503738765	0.772406631	5.561
ARN-509	−0.93691652	0.02300543	3.673
Target biomarker : TP53(*)			
Potential drug	**Regulation ability (L1000)**	**Sensitivity (PRISM)**	**Toxicity (LC50, mol/kg)**
Gemcitabine	−0.537988045	−2.417963872	2.381
Guanadrel	−0.915635131	−0.411395434	2.38
Bemegride	−3.718984186	1.008142951	1.532

Notes: *Denotes the mutation; +denotes overexpression on the corresponding biomarker.
Abbreviations: FOXO3: Forkhead box O3; LC50: Lethal concentration 50%; PRISMA: Pharmaceutical Regulatory Information System; TP53: Tumor suppressor p53.

target pairs into feature vectors to enable input into the DNN model. To generate the feature vectors, we used the Protein Feature Server and PyBioMed tool in a Python 3.7 environment. The drug features encompass widely utilized structural and physicochemical data, while the target features are derived from the structural and physicochemical characteristics of proteins and peptides, determined from their amino acid sequences. Each drug-target pair was combined into a single feature vector. The feature vector for the *i*-th drug-target pair in DTI databases can be presented as:

$$q^{i}_{drug-target} = \left[d_1, d_2, \ldots, d_{A-1}, d_A, t_1, t_2, \ldots, t_{B-1}, t_B\right]^i = \left[D, T\right]^i \quad \text{(XXXV)}$$

for i=1,2…,180315, $A + B = 1359$

The total feature vector dataset consists of 180,315 entries, including 80,291 experimentally validated DTIs and 100,024 unvalidated interactions. To address the imbalance in the dataset, we downsampled the unvalidated interactions to match the number of validated entries. Before training the DNN-DTI model, we standardized and transformed the drug-target interaction data because of variations in units among the different feature vectors. Standardization highlights the differences between each feature vector. The standardization of the features is shown as follows:

$$d^{*}_{a} = \frac{d_a - \mu_a}{\sigma_a} \quad for\ a = 1,2\ldots,A-1, A \quad \text{(XXXVI)}$$

where d_a represents the a-th drug feature, and d^{*}_{a} indicates the a-th drug feature after standardization; σ_a and μ_a refer to the standard deviation and mean of the a-th drug feature, respectively. A denotes the total number of drug features.

Similarly,

$$t^{*}_{b} = \frac{t_b - \mu_b}{\sigma_b} \quad for\ b = 1,2\ldots,B-1, B \quad \text{(XXXVII)}$$

where t_b denotes the b-th target feature, and t^{*}_{b} represents the b-th target feature after standardization; σ_b and μ_b refer to the standard deviation and mean of the b-th target feature, respectively; B represents the total number of target features.

Given that the DNN-based DTI model (Figure 3) requires 996 input nodes, the total (A+B) feature vector dimension needed to be reduced so that these drug-target feature vectors can be input to train the DNN-DTI model. By selecting the top 85% significant features for both drugs and targets using the principal component analysis (PCA), we reduced the dimensionality of the features from 1,359 to 996. This reduction aligns with the model's input layer dimension and enhances training performance.[38]

All the aforementioned data preprocessing steps were performed to enable the DNN-DTI model to effectively learn from feature data of drug-target interactions. After completing the data preprocessing, we split the dataset into

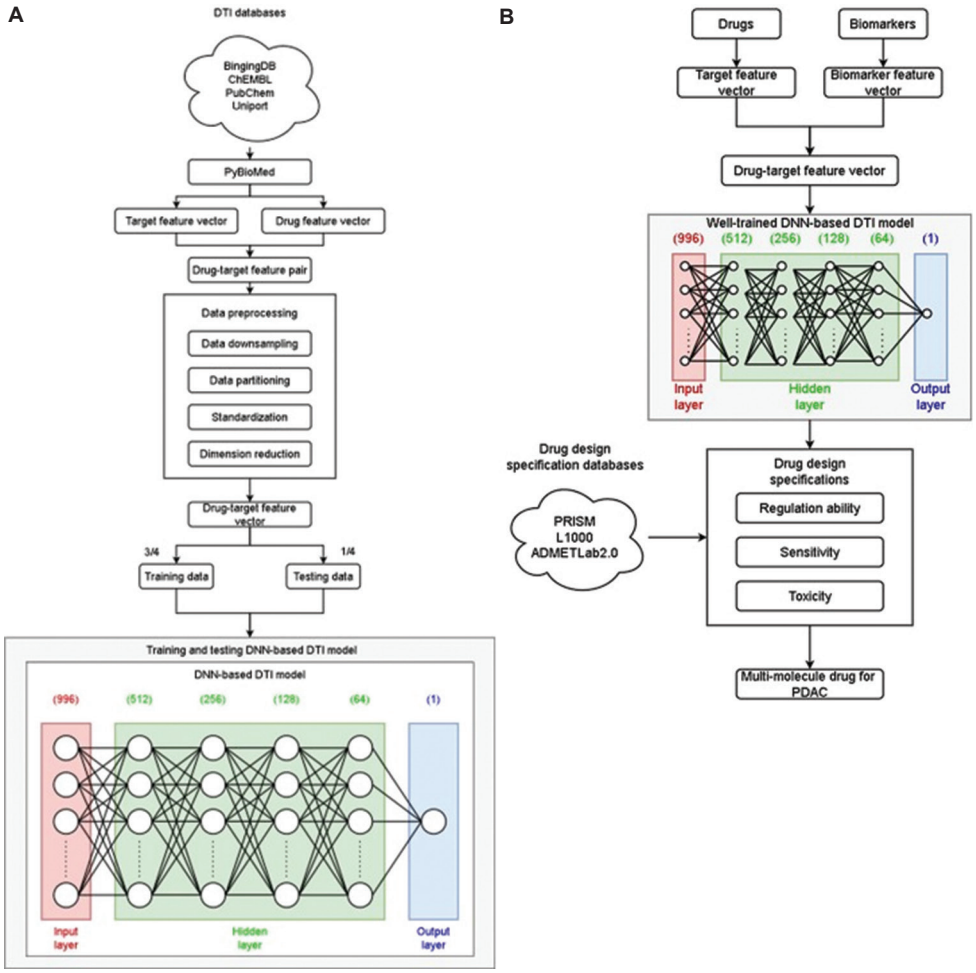


Figure 3. Flowchart of the systematic drug repurposing approach for treating PDAC. (A) Pre-training process of the deep neural network-based drug-target interaction model using drug-target interaction databases. (B) Prediction of candidate molecular drugs for PDAC biomarkers and selection of multi-molecular drug combination based on three drug design specifications
Abbreviations: DNN: Deep neural network; DTI: Drug-target interaction; PDAC: Pancreatic ductal adenocarcinoma.

75% for training and 25% for testing. We then utilized the DNN-DTI model to predict the probability of drug-target interactions. The architecture of the DNN model comprises an input layer with 996 neurons, followed by hidden layers with 512, 256, 128, and 64 neurons, respectively. The rectified linear unit (ReLU) activation function was applied in the hidden layers, and a dropout rate of 0.2 was set to prevent overfitting. The output layer consists of a single neuron with a sigmoid activation function, as drug-target interaction prediction is a binary classification problem. The mathematical formula for each layer is as follows:

$$O_n = A\left(w \times I_n + b\right) \quad \text{(XXXVIII)}$$

where A is the activation function (either ReLU or Sigmoid); I_n and O_n represent the input and output of the n-th feature vector, respectively; w and b are the weight vector and bias vector, respectively.

The output represents the probability of a drug-target interaction, ensuring the result lies within (0,1). The following binary cross-entropy was chosen as the loss function to address the binary classification problem of drug-target interactions:

$$C_n\left(p_n, \hat{p}_n\right) = -\left[p_n \log \hat{p}_n + \left(1 - p_n\right)\log\left(1 - \hat{p}_n\right)\right] \quad \text{(XXXIX)}$$

$$L\left(w, b\right) = \frac{1}{N}\sum_{n=1}^{N} C_n\left(p_n, \hat{p}_n\right) \quad \text{(XL)}$$

where n represents the nth sample, N is the total number of samples, p_n is the empirical probability of a positive interaction, $1-p_n$ is the empirical probability of a negative interaction, $\hat{p}_n$ is the estimated probability of a positive interaction, and $\hat{p}_n$ is the estimated probability of a negative interaction. The loss function computed the average loss $L(w, b)$ over all samples by averaging the loss $C_n\left(p_n, \hat{p}_n\right)$.

Finally, we used backpropagation and the Adam optimal learning algorithm with a learning rate of 0.001 to train the DNN-based DTI model. We set the number of epochs to 100 and the batch size to 100. The gradient update algorithm is given as follows:

$$\theta = \begin{bmatrix} w \\ b \end{bmatrix} \tag{XLI}$$

$$\theta^* = arg\min_{\theta} L(\theta) \tag{XLII}$$

$$\theta^{f-1} - \theta^f = \eta \nabla L\left(\theta^{f-1}\right) \tag{XLIII}$$

$$with\ \nabla L\left(\theta^{f-1}\right) = \begin{bmatrix} \frac{\partial L\left(\theta^{f-1}\right)}{\partial w} \\ \frac{\partial L\left(\theta^{f-1}\right)}{\partial b} \end{bmatrix}$$

where f represents the f-th iteration of the DNN training process, η is the learning rate, and ∇ denotes the gradient operator.

To evaluate the model, we employed five-fold cross-validation and used receiver operating characteristic (ROC) curves for the binary classification problem. The area under the ROC curve (AUC) is an important evaluation metric, where a higher AUC value indicates a better prediction of drug-target interaction. The formulas for calculating the ROC curves and AUC are as follows:

$$\text{TPR}\left(\text{True Positive Rate}\right) = \frac{TP}{TP + FN} \tag{XLIV}$$

$$\text{TNR}\left(\text{True Negative Rate}\right) = \frac{TN}{FP + TN} \tag{XLV}$$

$$\text{FPR}\left(\text{False Positive Rate}\right) = \frac{FP}{TN + FP} \tag{XLVI}$$

$$\text{FNR}\left(\text{False Negative Rate}\right) = \frac{FN}{FN + TP} \tag{XLVII}$$

where *TP* means the judgment is true and it is indeed true; *TN* means the judgment is false and it is indeed false; *FP* means the judgment is true, but it is actually false; *FN* means the judgment is false, but it is actually true.

Utilizing predictions from the DNN-based DTI model, we obtained three candidate molecular drugs for each key biomarker, as shown in Table 1. These drugs were screened based on some drug design specifications to identify potential molecular targets for PDAC. In this study, we considered pharmacological properties such as regulatory ability, sensitivity, and toxicity as key design specifications, and selected suitable candidate drugs from Table 1 based on these criteria. Specifically, we assessed the regulatory ability of drugs by referring to the LINCS L1000 level 5 dataset, which allowed us to identify drugs capable of regulating gene expression to normal levels. A regulatory ability value >0 indicates upregulation of gene expression, while a value <0 suggests downregulation. We also selected compounds with small absolute sensitivity values from the Pharmaceutical Regulatory Information System (PRISM) database to minimize interference with normal cells. Most importantly, in considering drug toxicity, we used tools from the ADMETlab 2.0 website and focused on LC50 values. LC50 represents the concentration at which 50% of organisms are lethally affected, and higher LC50 values indicate lower toxicity. Therefore, we selected compounds with higher LC50 values.

In conclusion, we followed three drug design specifications — toxicity, regulatory ability, and sensitivity — to screen potential molecular drugs for the biomarkers of PDAC, ultimately identifying two potential molecular drugs. These drugs were then combined into multi-target therapy for treating PDAC, as shown in Table 2. Ultimately, we successfully identified a suitable combination of molecular drugs for the treatment of PDAC.

3. Results

3.1. Overview of the systems biology approach to PDAC mechanisms and systematic drug repurposing and design

In this study, a systems biology approach was employed to investigate the carcinogenic mechanism of PDAC, utilizing big data mining and genome-wide microarray data. This approach led to the identification of crucial biomarkers of PDAC carcinogenesis, which were subsequently targeted for drug repurposing. A DNN-based DTI model was trained using DTI databases to predict potential drugs targeting these biomarkers. These molecular drugs were designed based on drug design specifications and their ability to restore the cellular functions of pancreatic cancer cells. Finally, the selected molecular drugs were considered as a multi-molecular therapeutic strategy of PDAC. The flowchart outlining the systems biology approach and drug repurposing process for PDAC is depicted in Figure 1.

Initially, to understand the carcinogenic mechanisms underlying PDAC and identify important biomarkers for therapeutic targeting, candidate GWGENs were constructed using a big database mining approach. The following databases were used: starBase,[27] DIP,[18] CircuitDB,[25] BioGRID,[20] IntAct,[19] HTRIdatabase,[22] ITFP,[23] MINT,[21] TRANSFAC,[24] and TargetScanHuman.[26] The candidate GWGENs are represented as a Boolean matrix,

Table 2. Filtering potential molecular drugs for pancreatic ductal adenocarcinoma based on three drug design specifications: regulatory ability, sensitivity, and toxicity, derived from the candidate molecular drugs listed in Table 1

Target drug	c-MYC (+)	FOXO3 (+)	TP53 (*)	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
MK-2206		↓		0.772406631	5.561
Gemcitabine	↓		✓	2.417963872	2.381

Notes: ✓ denotes drug-target interaction; ↓ denotes downregulation of the biomarker by the molecular drug.
Abbreviations: FOXO3: Forkhead box O3; LC50: Lethal concentration 50%; PRISMA: Pharmaceutical Regulatory Information System; TP53: Tumor suppressor p53.

where interactions between proteins and genes were encoded as 1 (interaction) and 0 (no interaction). The nodes in the GWGENs include proteins, genes, miRNAs, and lncRNAs.

Next, using systematic identification and order detection methods, false positives were removed from the candidate GWGENs based on whole-genome microarray data GSE183795, resulting in the construction of the real GWGEN for both PDAC and healthy controls, as shown in Figure S1. Although the number of nodes was reduced after the false positives were filtered out from candidate GWGENs, the real GWGENs for both PDAC and healthy controls remained too complex for direct annotation using KEGG pathway analysis. To simplify this, the number of nodes in GWGENs was reduced to 6,000 using the PNP method. The 6,000 significant nodes from the real GWGENs of PDAC and healthy controls were extracted, forming the core GWGENs for PDAC and healthy controls, as depicted in Figure S2.

For KEGG pathway enrichment analysis to interpret the carcinogenic mechanisms of PDAC, the DAVID functional annotation tool was employed, supported by references from PDAC-related literature. Based on the KEGG pathway annotations, core signaling pathways for PDAC and healthy controls were established, allowing the investigation of the mechanisms involved in PDAC carcinogenesis, as illustrated in Figure 2.

Based on these core signaling pathways, key biomarkers of PDAC carcinogenesis were identified as potential drug targets, which were implicated in downstream cellular dysfunctions associated with PDAC. Using the DNN-based DTI model trained with data from the DTI databases, the systematic drug repurposing and design process for PDAC therapy was carried out, as shown in Figure 3. The DTI databases served as the training set for the DNN model, which predicted candidate molecular drugs for PDAC drug targets. These candidate drugs were screened according to drug design specifications, including regulatory ability, sensitivity, and toxicity, to identify a suitable combination of molecular drugs for PDAC treatment.

3.2. Core signaling pathways of carcinogenic mechanisms of PDAC

Tumors are more likely to accelerate their development in an appropriate microenvironment. The microenvironment of PDAC involves complex interactions between tissues, blood vessels, immune cells, cytokines, and other molecules surrounding pancreatic cancer cells.[39] The characteristics of this microenvironment include fibrosis, immune cell infiltration, angiogenesis, and the presence of various cytokines and growth factors.[40,41] To identify relevant signaling pathways based on their cellular functions, we consulted existing research literature. Key pathways involved in pancreatic fibrosis include transforming growth factor β, Wnt, and phosphoinositide 3-kinase (PI3K)-protein kinase B (AKT) signaling pathways.[42] Pathways associated with immune cell infiltration include Janus kinase-signal transducer and activator of transcription (STAT) pathway, PI3K-AKT, and mammalian target of rapamycin (mTOR) signaling pathways.[43] Additionally, the mitogen-activated protein kinase, PI3K-AKT, and mTOR signaling pathways have been implicated in the creation of new blood vessels (angiogenesis).[44] These six signaling pathways provide a comprehensive framework for investigating the oncogenic mechanisms of PDAC.

Through KEGG enrichment pathway analysis and annotation of the core GWGENs of PDAC and healthy controls, we identified the core signaling pathways involved in PDAC carcinogenesis, as shown in Figure 2. By examining the core signaling pathways and their downstream target genes, we explored the carcinogenic mechanisms of PDAC and identified key biomarkers as drug targets. This approach ultimately aided in the discovery of multi-molecular drugs for the treatment of PDAC.

3.3. Core signaling pathways in healthy controls

In healthy controls, the core signaling pathways primarily include PI3K-AKT, TP53, and interleukin 17 (IL-17) signaling pathways. The TP53 pathway plays a crucial role in regulating the cell cycle by modulating the expression and activity of genes such as *p21*, *CDK4/6*, *CDK2*, *BAX*, *E2F*, and *MUVB*. These interactions influence cell

proliferation and survival. The IL-17 signaling pathway involves proteins such as activator protein 1 (AP-1) and tumor necrosis factor receptor-associated factor 6 (TRAF6), which work together to regulate the expression of downstream genes. Additionally, tumor necrosis factor α (TNF-α) is a key cytokine involved in immune responses and inflammation regulation. Signaling through the IL-17 pathway often triggers the production of *TNFα*, thereby impacting the inflammatory responses. The activation of the IL-17 signaling pathway is regulated by genes including *IL17B*, *IL17RB*, *TRAF6*, AP-1, and *TNFα*.

Through KEGG pathway enrichment analysis and annotation of the core GWGENs for both PDAC and healthy controls, as shown in Figure 2, we obtained the core signaling pathways and their downstream target genes. This analysis provides a deeper understanding of the carcinogenic mechanisms of PDAC. By comparing the core signaling pathways and their downstream target genes between PDAC and healthy controls, we identified significant biomarkers as drug targets, thus facilitating the discovery of more effective multi-molecular drugs for treating PDAC. Furthermore, recent studies have suggested that miRNAs may be the answer to cancer treatment. Based on the core GWGENs of healthy controls, we identified miRNA LET-7C. According to recent research, miRNA LET-7C regulates key genes such as *Ras, STAT1/3*, and *c-MYC* in the core signaling pathways of PDAC.[45]

3.4. Selection of drug targets by investigating the pathogenesis of PDAC and systematic drug repurposing via deep neural network-based drug-target interactions model

Based on the core signaling pathways of PDAC patients and the abnormal downstream cellular dysfunctions compared with the healthy controls in Figure 2, we selected key biomarkers, such as c-MYC, FOXO3, and TP53, as drug targets. Our aim is to identify potential multi-molecular drugs that could restore the abnormal expression of these key biomarkers to normal levels for the treatment of PDAC.

c-MYC, a TF located downstream of the MAPK signaling pathway, is important in regulating cell growth, proliferation, and apoptosis.[46] Research indicates that overexpression of c-MYC is strongly associated with the development of PDAC and plays a significant role in its early transformation.[47] Excessive activation of c-MYC promotes the proliferation of PDAC cells, thereby accelerating tumor growth.[48] Moreover, c-MYC inhibits apoptosis in PDAC cells, further enhancing tumor progression.[49] Thus, downregulating c-MYC can effectively inhibit PDAC cell proliferation and survival, offering therapeutic potential.[50] Additionally, c-MYC contributes to PDAC resistance to chemotherapy. Research has demonstrated that overexpression of MYC enhances the resistance of PDAC cells to chemotherapy,[51] thereby reducing treatment efficacy. By regulating c-MYC to restore cellular function, the sensitivity of PDAC cells to existing treatments can be enhanced, thereby improving therapeutic outcomes.

Forkhead box O3 is a TF that plays a crucial role in regulating cell proliferation, apoptosis, and metabolism.[52] It has been shown to significantly involve in the development and progression of PDAC.[53] Dysregulation of FOXO3 is closely linked to the onset and progression of PDAC, with its expression in PDAC being much higher than in normal tissues, although the precise mechanism remains unclear.[54] Scholars speculate that FOXO3 may be upregulated or degraded in cancer tissues, leading to its accumulation in the nucleus and increased expression levels. In normal tissues, FOXO3 is transcribed at low levels and rapidly degraded.[55] In our study, FOXO3 regulates several target genes, and our goal is to restore FOXO3 to normal expression levels to mitigate downstream cellular dysfunction.

The *TP53* gene encodes the tumor suppressor protein p53, one of the most commonly mutated genes in PDAC.[56] TP53 regulates the expression of many genes involved in cell cycle control, apoptosis, and senescence, as well as influencing the immune microenvironment of PDAC. Inactivation of the *TP53* gene disrupts these processes, leading to altered immune responses.[56] Research indicates that the loss of *TP53* expression allows cells harboring the *KRASG12D* mutation to survive, promoting tumor formation and metastasis.[57] Our goal is to restore cellular function to normal by targeting mutated TP53, thereby achieving therapeutic effects.

After selecting these important biomarkers of PDAC carcinogenesis as potential drug targets, we obtained DTI data from databases such as KEGG,[31] BIDD,[32] UniProt,[33] DrugBank,[34] PubChem,[35] ChEMBL,[36] and STITCH.[37] Using these data, we trained a DNN-based DTI model to predict molecular drugs for these targets. The training dataset consists of 180,315 drug-target interactions, including 80,291 experimentally verified drug-target interactions and 100,024 unverified interactions. To address the potential imbalance between the two categories in the training dataset, we randomly selected an equal number of drug-target interactions from each database. Before training the DNN-DTI model, the interaction data were transformed and standardized, followed by dimensionality reduction using PCA, which reduced the transformed features from 1,359 to 996. This step is necessary to match the input layer dimensions of the model. As shown in Figure 3, the input layer of the DNN-DTI model consists of 996 nodes, followed by four hidden layers of

512, 256, 128, and 64 nodes, respectively, each using the ReLU activation function. The output layer consists of a single node with sigmoid activation function. To prevent overfitting during training, we added a dropout layer. After training the DNN-DTI model to predict candidate molecular drugs for these drug targets, we evaluated the model's learning effectiveness. Figures S3 and S4 show the accuracy and loss during the training process, respectively. The five-fold cross-validation was employed to assess the model's performance, achieving an average accuracy of 98.3% and a standard deviation of 0.138%, as shown in Table S1. Additionally, the AUC was used to evaluate the model's classification performance, as shown in Figure S5. An AUC of 0.5 represents random guessing, while an AUC of 1 indicates perfect classification. The DNN-DTI model achieved an AUC of 0.980, indicating that its predictive ability is much superior to random guessing and close to perfect classification. This highly efficient DNN-based DTI model enables us to accurately predict the probability of interactions between drugs and the selected biomarkers.

To identify suitable potential drugs, we considered three drug design specifications to ensure the rationality and effectiveness of the candidate multi-molecular drugs predicted by the DNN-DTI model. These specifications include regulatory ability, sensitivity, and toxicity, among other pharmacological properties. The regulatory ability of the drugs was assessed using the LINCS L1000 level 5 dataset, guiding the selection of drugs that could restore key biomarkers to their normal expression levels.[58] A regulatory ability value >0 indicates an increase in gene expression, while a value <0 indicates a decrease. Additionally, the sensitivity of the drugs was assessed using the PRISM dataset,[59] and we selected drugs with small absolute sensitivity values to avoid excessive chemical perturbation from potential drugs. Finally, we evaluated the drug toxicity using tools from the ADMETlab 2.0 website, which calculates LC50.[60] A higher LC50 value indicates lower toxicity, which helps avoid life-threatening from low drug doses. Table 1 presents the candidate molecular drugs predicted by the DNN-DTI model for the selected biomarkers, listing their relevant pharmacological properties, such as regulatory ability, sensitivity, and toxicity. Based on these drug design specifications, we selected two potential molecular drugs, MK-2206 and gemcitabine, which have adequate regulatory ability, normal sensitivity (small absolute value), and weaker toxicity, as shown in Table 2. These drugs were combined into a multi-molecular therapeutic strategy for PDAC.

4. Discussion

Currently, treatment options for exocrine ductal adenocarcinoma, also known as PDAC, include surgical resection, radiotherapy, chemotherapy, immunotherapy, and targeted therapy. However, not all patients are suitable candidates for surgery, and standard chemotherapy regimens may be ineffective due to the complex interactions between healthy pancreatic cells, cancer cells, and the tumor microenvironment, leading to drug resistance and suboptimal therapeutic outcomes. Therefore, there is a pressing need to explore new treatment approaches.

Given the substantial financial costs and time required for the discovery of new drugs, repurposing existing drugs for new therapeutic indications has become an attractive alternative. Drug repurposing involves identifying new uses for already-approved drugs beyond their original medical indications.[17]

In this study, we employed database mining and genome-wide microarray data from PDAC and healthy controls, utilizing systems biology approaches to identify their core GWGENs. These networks were annotated using the KEGG database to establish the core signaling pathways of PDAC and the associated downstream cellular dysfunctions, as shown in Figure 2. After investigating the oncogenic mechanisms of PDAC and identifying key biomarkers suitable for drug targeting, we trained a DNN-based DTI model using data from DTI databases to predict the probability of interaction between these biomarkers and candidate molecular drugs. The model was validated using five-fold cross-validation, as shown in Table S1, and the drug repurposing flowchart is presented in Figure 3. We subsequently screened potential drugs as a multi-molecular therapy based on drug design specifications, focusing on sensitivity, toxicity, and regulatory capability. Ultimately, we predicted a combination of potential molecular drugs, MK-2206 and gemcitabine, to modulate the overexpression of c-MYC and FOXO3, as well as the mutation of TP53, as shown in Table 2.

Gemcitabine is one of the most widely used treatments for PDAC. As a deoxycytidine nucleoside analog,[10] it inhibits DNA chain elongation by phosphorylation after entering the cells, leading to cell apoptosis and death. However, the therapeutic effect of gemcitabine is limited by its unstable metabolism and the potential for drug resistance.[61] As a result, it is often used in combination with other drugs, such as Fluorouracil and paclitaxel,[62,63] to enhance treatment efficacy. Studies have pointed out that gemcitabine can induce the activation of mutated p53, leading to cancer cell death,[64] as the drug causes the accumulation of *Bax* downstream of TP53, which induces apoptosis.[65,66] In our research, we referenced the LINCS L1000 level 5 dataset and found that gemcitabine can reduce the expression c-MYC.[67] However, the high expression of c-MYC may induce resistance to gemcitabine.[68] Therefore,

Table 3. Overview of potential multi-molecular drugs screened for pancreatic ductal adenocarcinoma treatment based on three drug design specifications using a systematic drug discovery approach

Target drug	c-MYC	FOXO3	TP53	Sensitivity (PRISM)	Toxicity (LC50, mol/kg)
MK-2206		•		0.772406631	5.561
Gemcitabine	•		•	2.417963872	2.381
MK-2206			Gemcitabine		

Note: • Denotes drug targeting on the corresponding biomarker.
Abbreviations: FOXO3: Forkhead box O3; LC50: Lethal concentration 50%; PRISMA: Pharmaceutical Regulatory Information System; TP53: Tumor suppressor p53.

we selected gemcitabine as a potential drug for further evaluation.

MK-2206 is an AKT inhibitor. Abnormal activation of the PI3K/AKT signaling pathway affects cell metabolism and survival, cell cycle progression, apoptosis regulation, protein synthesis, and genome instability.[69] Since the PI3K/AKT pathway is frequently dysregulated in many cancers, MK-2206 inhibits AKT and downregulates FOXO3,[70,71] thereby exhibiting anti-tumor activity and increasing cancer cell sensitivity to chemotherapy and radiotherapy. MK-2206 is often used in combination with other chemotherapy drugs to improve therapeutic efficacy.[72] Current studies show that MK-2206 is well tolerated and can effectively block AKT signaling,[73] making it a promising candidate for repurposing as a potential drug for PDAC treatment.

However, this study has several limitations. First, the analysis was based on GSE183795 microarray data, and the quality and completeness of this dataset directly influence the accuracy of the results. If the dataset contains significant noise or incomplete data, it may lead to erroneous conclusions. Second, the findings of this study are primarily based on mathematical modeling and bioinformatics analysis, and lack direct experimental validation. Predictions made without experimental confirmation may face challenges in practical application, highlighting the need for further laboratory experiments to confirm the results. While there is limited experimental research on the combination treatment of MK-2206 and gemcitabine, existing studies suggest that this combination may be effective.[74,75] Additionally, although our study predicts multi-drug combination, the actual pharmacodynamic interactions and effects of this combination need to be verified through *in vivo* and *in vitro* experiments. It is also important to consider potential synergistic or antagonistic effects between the drugs, which may not have been fully captured in the model predictions.

These limitations should be addressed in future research to enhance the reliability and applicability of drug repurposing strategies in PDAC treatment.

5. Conclusion

In this research, we employed a systems biology approach to construct candidate GWGENs by mining large-scale databases. We then applied system identification and order detection methods to eliminate false positive interactions within candidate GWGENs, thereby generating the real GWGENs for PDAC and healthy controls. Using the principal network projection method, we extracted the core GWGENs for both PDAC and healthy controls. To elucidate the oncogenic mechanisms of PDAC, we annotated the core GWGENs with KEGG pathways to identify core signaling pathways in PDAC and healthy controls. By comparing the core signal pathways between PDAC and healthy controls, we investigated the association between these pathways and downstream cellular dysfunctions in PDAC, identifying key biomarkers involved in the carcinogenesis of PDAC as potential drug targets. With these drug targets, we utilized a DNN-based DTI model to predict the probability of interaction between the drugs and their targets. Based on the drug's regulatory ability, sensitivity, and toxicity, we selected a potential multi-molecular drug combination. Ultimately, we identified MK-2206 and gemcitabine as a promising combination for PDAC treatment, targeting significant biomarkers such as c-MYC, FOXO3, and TP53, as shown in Table 3. While the Drug Interaction Checker suggests that there is no known interaction between these

two drugs, their interaction in clinical treatment for PDAC still requires further clinical and experimental verification.

Conclusion

In conclusion, we integrated systems biology methods with deep learning techniques to systematically repurpose existing drugs for PDAC therapy. By utilizing systems biology to identify drug targets through large-scale databases and genome-wide microarray data from both PDAC and non-PDAC samples, and applying deep learning for the DTI model to identify potential molecular drugs based on three drug design specifications, we have developed a systematic approach to accelerate drug development for PDAC. We hope that this drug repurposing strategy can be applied to other diseases in the future, further advancing the field of systems drug development.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: All authors
Formal analysis: All authors
Investigation: Yi-Hsin Tsai
Methodology: All authors
Writing – original draft: Yi-Hsin Tsai
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The RNA-sequencing datasets from PDAC patients and healthy controls can be obtained from GSE183795 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183795, accessed on 10 March 2023).

References

1. Ma Y, Wu Q, Li X, Gu X, Xu J, Yang J. Pancreatic cancer: From bench to bedside. *Ann Transl Med*. 2016;4(23):458.

   doi: 10.21037/atm.2016.11.57

2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. 2014;74(11):2913-2921.

   doi: 10.1158/0008-5472.Can-14-0155

3. Rahib L, Wehner MR, Matrisian LM, Nead KT. estimated projection of US cancer incidence and death to 2040. *JAMA Netw Open*. 2021;4(4):e214708.

   doi: 10.1001/jamanetworkopen.2021.4708

4. Strobel O, Lorenz P, Hinz U, *et al*. Actual five-year survival after upfront resection for pancreatic ductal adenocarcinoma: Who beats the odds? *Ann Surg*. 2022;275(5):962-971.

   doi: 10.1097/sla.0000000000004147

5. Sung H, Ferlay J, Siegel RL, *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(3):209-249.

   doi: 10.3322/caac.21660

6. Lynch SM, Vrieling A, Lubin JH, *et al.* Cigarette smoking and pancreatic cancer: A pooled analysis from the pancreatic cancer cohort consortium. *Am J Epidemiol*. 2009;170(4):403-413.

   doi: 10.1093/aje/kwp134

7. Mela A, Rdzanek E, Tysarowski A, *et al*. The impact of changing the funding model for genetic diagnostics and improved access to personalized medicine in oncology. *Expert Rev Pharmacoecon Outcomes Res*. 2023;23(1):43-54.

   doi: 10.1080/14737167.2023.2140139

8. Park W, Chawla A, O'Reilly EM. Pancreatic cancer: A review. *JAMA*. 2021;326(9):851-862.

   doi: 10.1001/jama.2021.13027

9. Elsayed M, Abdelrahim M. The latest advancement in pancreatic ductal adenocarcinoma therapy: A review article for the latest guidelines and novel therapies. *Biomedicines*. 2021;9(4):389.

   doi: 10.3390/biomedicines9040389

10. Zeng S, Pöttler M, Lan B, Grützmann R, Pilarsky C, Yang H. Chemoresistance in pancreatic cancer. *Int J Mol Sci*. 2019;20(18):4504.

    doi: 10.3390/ijms20184504

11. Hughes J, Rees S, Kalindjian S, Philpott K. Principles of early drug discovery. *Br J Pharmacol*. 2011;162(6):1239-1249.

    doi: 10.1111/j.1476-5381.2010.01127.x

12. Weaver MF, Hopper JA, Gunderson EW. Designer drugs 2015: Assessment and management. *Addict Sci Clin Pract*. 2015;10(1):8.

    doi: 10.1186/s13722-015-0024-7

13. Zhan YP, Chen BS. Drug target identification and drug repurposing in psoriasis through systems biology approach, DNN-based DTI model and genome-wide microarray data. *Int J Mol Sci*. 2023;24(12):10033.

doi: 10.3390/ijms241210033

14. Wang CT, Chen BS. Drug discovery for periodontitis treatment based on big data mining, systems biology, and deep learning methods. *SynBio*. 2023;1(1):116-143.

doi: 10.3390/synbio1010009

15. Hsu BW, Chen BS. Genetic and epigenetic host-virus network to investigate pathogenesis and identify biomarkers for drug repurposing of human respiratory syncytial virus via real-world two-side RNA-Seq data: Systems biology and deep-learning approach. *Biomedicines*. 2023;11(6):1531.

doi: 10.3390/biomedicines11061531

16. Lin YC, Chen BS. Identifying drug targets of oral squamous cell carcinoma through a systems biology method and genome-wide microarray data for drug discovery by deep learning and drug design specifications. *Int J Mol Sci*. 2022;23(18):10409.

doi: 10.3390/ijms231810409

17. Kulkarni VS, Alagarsamy V, Solomon VR, Jose PA, Murugesan S. Drug repurposing: An effective tool in modern drug discovery. *Russ J Bioorg Chem*. 2023;49(2):157-166.

doi: 10.1134/s1068162023020139

18. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The database of interacting proteins: 2004 update. *Nucleic Acids Res*. 2004;32(suppl_1):D449-D451.

doi: 10.1093/nar/gkh086

19. Orchard S, Ammari M, Aranda B, *et al*. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 2013;42(D1):D358-D363.

doi: 10.1093/nar/gkt1115

20. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res*. 2006;34(suppl_1):D535-D539.

doi: 10.1093/nar/gkj109

21. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. MINT: A molecular INTeraction database. *FEBS Lett*. 2002;513(1):135-140.

doi: 10.1016/S0014-5793(01)03293-8

22. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*. 2012;13(1):405.

doi: 10.1186/1471-2164-13-405

23. Zheng G, Tu K, Yang Q, *et al*. ITFP: An integrated platform of mammalian transcription factors. *Bioinformatics*. 2008;24(20):2416-2417.

doi: 10.1093/bioinformatics/btn439

24. Wingender E, Chen X, Hehl R, *et al*. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res*. 2000;28(1):316-319.

doi: 10.1093/nar/28.1.316

25. Friard O, Re A, Taverna D, De Bortoli M, Corá D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*. 2010;11(1):435.

doi: 10.1186/1471-2105-11-435

26. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4:e05005.

doi: 10.7554/eLife.05005

27. Li JH, Liu S, Zhou H, Qu LH, Yang JH. StarBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2013;42(D1):D92-D97.

doi: 10.1093/nar/gkt1248

28. Chen BS, Wu CC. Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering. *Cells*. 2013;2(4):635-688.

doi: 10.3390/cells2040635

29. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. *Dordrecht Netherlands D Reidel*. 1986;81(10.5555):26853.

30. Chang S, Wang LHC, Chen BS. Investigating core signaling pathways of hepatitis b virus pathogenesis for biomarkers identification and drug discovery via systems biology and deep learning method. *Biomedicines*. 2020;8(9):320.

doi: 10.3390/biomedicines8090320

31. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.

doi: 10.1093/nar/28.1.27

32. Wang Y, Zhang S, Li F, *et al*. Therapeutic target database 2020: Enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res*. 2019;48(D1):D1031-D1041.

doi: 10.1093/nar/gkz981

33. Consortium TU. UniProt: A hub for protein information. *Nucleic Acids Res*. 2014;43(D1):D204-D212.

doi: 10.1093/nar/gku989

34. Knox C, Law V, Jewison T, *et al*. DrugBank 3.0: A comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res*. 2010;39(suppl_1):D1035-D1041.

doi: 10.1093/nar/gkq1126

35. Kim S, Thiessen PA, Bolton EE, *et al*. PubChem substance and compound databases. *Nucleic Acids Res*. 2015;44(D1):D1202-D1213.

doi: 10.1093/nar/gkv951

36. Gaulton A, Hersey A, Nowotka M, *et al*. The ChEMBL database in 2017. *Nucleic Acids Res*. 2016;45(D1):D945-D954.

doi: 10.1093/nar/gkw1074

37. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res*. 2007;36(suppl_1):D684-D688.

doi: 10.1093/nar/gkm795

38. Ringnér M. What is principal component analysis? *Nat Biotechnol*. 2008;26(3):303-304.

doi: 10.1038/nbt0308-303

39. Huber M, Brehm CU, Gress TM, *et al*. The immune microenvironment in pancreatic cancer. *Int J Mol Sci*. 2020;21(19):7307.

doi: 10.3390/ijms21197307

40. Truong LH, Pauklin S. Pancreatic cancer microenvironment and cellular composition: Current understandings and therapeutic approaches. *Cancers* (*Basel*). 2021;13(19):5028.

doi: 10.3390/cancers13195028

41. Cortesi M, Zanoni M, Pirini F, *et al*. Pancreatic cancer and cellular senescence: Tumor microenvironment under the spotlight. *Int J Mol Sci*. 2022;23(1):254.

doi: 10.3390/ijms23010254

42. Arneth B. Tumor microenvironment. *Medicina* (*Kaunas*). 2020;56(1):15.

doi: 10.3390/medicina56010015

43. Stopa KB, Kusiak AA, Szopa MD, Ferdek PE, Jakubowska MA. Pancreatic cancer and its microenvironment-recent advances and current controversies. *Int J Mol Sci*. 2020;21(9):3218.

doi: 10.3390/ijms21093218

44. Javadrashid D, Baghbanzadeh A, Derakhshani A, *et al*. Pancreatic cancer signaling pathways, genetic alterations, and tumor microenvironment: The barriers affecting the method of treatment. *Biomedicines*. 2021;9(4):373.

doi: 10.3390/biomedicines9040373

45. Ma Y, Shen N, Wicha MS, Luo M. The roles of the let-7 family of MicroRNAs in the regulation of cancer stemness. *Cells*. 2021;10(9):2415.

doi: 10.3390/cells10092415

46. Hessmann E, Schneider G, Ellenrieder V, Siveke JT. MYC in pancreatic cancer: Novel mechanistic insights and their translation into therapeutic strategies. *Oncogene*. 2016;35(13):1609-1618.

doi: 10.1038/onc.2015.216

47. Wolfer A, Ramaswamy S. MYC and metastasis. *Cancer Res*. 2011;71(6):2034-2037.

doi: 10.1158/0008-5472.Can-10-3776

48. Liu YH, Gan YX, Chen JZ, Jiang YX, Huang F, Tang D. Targeting MYC to combat pancreatic cancer. *J Explor Rese Pharmacol*. 2022;7(3):164-178.

doi: 10.14218/jerp.2021.00015

49. Wirth M, Mahboobi S, Krämer OH, Schneider G. Concepts to target MYC in pancreatic cancer. *Mol Cancer Ther*. 2016;15(8):1792-1798.

doi: 10.1158/1535-7163.Mct-16-0050

50. Ala M. Target c-Myc to treat pancreatic cancer. *Cancer Biol Ther*. 2022;23(1):34-50.

doi: 10.1080/15384047.2021.2017223

51. Farrell AS, Joly MM, Allen-Petersen BL, *et al*. MYC regulates ductal-neuroendocrine lineage plasticity in pancreatic ductal adenocarcinoma associated with poor outcome and chemoresistance. *Nature Commun*. 2017;8(1):1728.

doi: 10.1038/s41467-017-01967-6

52. Fujiwara-Tani R, Sasaki T, Bhawal UK, *et al*. Nuclear MAST4 suppresses FOXO3 through interaction with AKT3 and induces chemoresistance in pancreatic ductal carcinoma. *Int J Mol Sci*. 2024;25(7):4056.

doi: 10.3390/ijms25074056

53. Luo X, Yang Z, Liu X, *et al*. The clinicopathological significance of forkhead box P1 and forkhead box O3a in pancreatic ductal adenocarcinomas. *Tumor Biol*. 2017;39(5):1010428317699129.

doi: 10.1177/1010428317699129

54. Feng S, Jiang ZJ, Yu D, Li J, Liu G, Sun JJ. FOXO3a expression and its diagnostic value in pancreatic ductal adenocarcinoma. *Int J Clin Exp Pathol*. 2018;11(11):5422-5429.

55. Yu S, Yu Y, Sun Y, *et al*. Activation of FOXO3a suggests good prognosis of patients with radically resected gastric cancer. *Int J Clin Exp Pathol*. 2015;8(3):2963-2970.

56. McCubrey JA, Yang LV, Abrams SL, *et al*. Effects of TP53 mutations and mirs on immune responses in the tumor microenvironment important in pancreatic cancer progression. *Cells*. 2022;11(14):2155.

doi: 10.3390/cells11142155

57. Stefanoudakis D, Frountzas M, Schizas D, Michalopoulos NV, Drakaki A, Toutouzas KG. Significance of TP53, CDKN2A, SMAD4 and KRAS in pancreatic cancer. *Curr Issues Mol Biol*. 2024;46(4):2827-2844.

doi: 10.3390/cimb46040177

58. Subramanian A, Narayan R, Corsello SM, *et al*. A next

generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437-1452.e17.

doi: 10.1016/j.cell.2017.10.049

59. Corsello SM, Nagari RT, Spangler RD, *et al*. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer*. 2020;1(2):235-248.

doi: 10.1038/s43018-019-0018-6

60. Xiong G, Wu Z, Yi J, *et al*. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res*. 2021;49(W1):W5-W14.

doi: 10.1093/nar/gkab255

61. Amrutkar M, Gladhaug IP. Pancreatic cancer chemoresistance to gemcitabine. *Cancers* (*Basel*). 2017;9(11):157.

doi: 10.3390/cancers9110157

62. Adamska A, Domenichini A, Falasca M. Pancreatic ductal adenocarcinoma: Current and evolving therapies. *Int J Mol Sci*. 2017;18(7):1338.

doi: 10.3390/ijms18071338

63. Coppola A, Farolfi T, La Vaccara V, *et al*. Neoadjuvant treatments for pancreatic ductal adenocarcinoma: Where we are and where we are going. *J Clin Med*. 2023;12(11):3677.

doi: 10.3390/jcm12113677

64. Kielb SJ, Shah NL, Rubin MA, Sanda MG. Functional p53 mutation as a molecular determinant of paclitaxel and gemcitabine susceptibility in human bladder cancer. *J Urol*. 2001;166(2):482-487.

65. Hill R, Rabb M, Madureira PA, *et al*. Gemcitabine-mediated tumour regression and p53-dependent gene expression: Implications for colon and pancreatic cancer therapy. *Cell Death Dis*. 2013;4(9):e791.

doi: 10.1038/cddis.2013.307

66. Achanta G, Pelicano H, Feng L, Plunkett W, Huang P. Interaction of p53 and DNA-PK in response to nucleoside analogues: Potential role as a sensor complex for DNA damage. *Cancer Res*. 2001;61(24):8723-8729.

67. Cao P, Zhang W, Qiu J, Tang Z, Xue X, Feng T. Gemcitabine inhibits the progression of pancreatic cancer by restraining the WTAP/MYC chain in an m6A-dependent manner. *Cancer Res Treat*. 2024;56(1):259-271.

doi: 10.4143/crt.2022.1600

68. Yao J, Huang M, Shen Q, *et al*. c-Myc-PD-L1 axis sustained gemcitabine-resistance in pancreatic cancer. *Front Pharmacol*. 2022;13:851512.

doi: 10.3389/fphar.2022.851512

69. Ebrahimi S, Hosseini M, Shahidsales S, *et al*. Targeting the Akt/PI3K signaling pathway as a potential therapeutic strategy for the treatment of pancreatic cancer. *Curr Med Chem*. 2017;24(13):1321-1331.

doi: 10.2174/0929867324666170206142658

70. Konopleva MY, Walter RB, Faderl SH, *et al*. Preclinical and early clinical evaluation of the oral AKT inhibitor, MK-2206, for the treatment of acute myelogenous leukemia. *Clin Cancer Res*. 2014;20(8):2226-2235.

doi: 10.1158/1078-0432.Ccr-13-1978

71. Simioni C, Martelli AM, Cani A, *et al*. The AKT inhibitor MK-2206 is cytotoxic in hepatocarcinoma cells displaying hyperphosphorylated AKT-1 and synergizes with conventional chemotherapy. *Oncotarget*. 2013;4(9):1496-1506.

doi: 10.18632/oncotarget.1236

72. Hirai H, Sootome H, Nakatsuru Y, *et al*. MK-2206, an allosteric akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs *in vitro* and *in vivo*. *Mol Cancer Ther*. 2010;9(7):1956-1967.

doi: 10.1158/1535-7163.Mct-09-1012

73. Yap TA, Yan L, Patnaik A, *et al*. First-in-man clinical trial of the oral pan-AKT inhibitor MK-2206 in patients with advanced solid tumors. *J Clin Oncol*. 2011;29(35):4688-4695.

doi: 10.1200/jco.2011.35.5263

74. Wang Z, Luo G, Qiu Z. Akt inhibitor MK-2206 reduces pancreatic cancer cell viability and increases the efficacy of gemcitabine. *Oncol Lett*. 2020;19(3):1999-2004.

doi: 10.3892/ol.2020.11300

75. Elnaggar MS, Sminia P, Shehata S, Fedrigo C, Peters G. 1676P - gemcitabine, PI3kinase-Akt pathway inhibition and radiation in human glioma cell lines. *Ann Oncol*. 2012;23:ix537.

doi: 10.1016/S0923-7534(20)34222-8