



OPEN ACCESS

EDITED BY

HaiHui Huang,
Shaoguan University, China

REVIEWED BY

Xi Long,
Hong Kong University of Science and
Technology, Hong Kong SAR, China
Eko Mugiyanto,
Muhammadiyah University of Pekajangan
Pekalongan, Indonesia

*CORRESPONDENCE

Adrian Gu,
✉ adrgu666@gmail.com

RECEIVED 19 June 2025

ACCEPTED 15 September 2025

PUBLISHED 29 September 2025

CITATION

Gu A and Chen JY (2025) GETgene-AI: a
framework for prioritizing actionable cancer
drug targets.
Front. Syst. Biol. 5:1649758.
doi: 10.3389/fsysb.2025.1649758

COPYRIGHT

© 2025 Gu and Chen. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

GETgene-AI: a framework for prioritizing actionable cancer drug targets

Adrian Gu^{1*} and Jake Y. Chen^{1,2}

¹AlphaMind Club, Birmingham, AL, United States, ²Systems Pharmacology AI Research Center, School of Medicine, The University of Alabama at Birmingham, Birmingham, AL, United States

Prioritizing actionable drug targets is a critical challenge in cancer research, where high-dimensional genomic data and the complexity of tumor biology often hinder effective prioritization. To address this, we developed GETgene-AI, a novel computational framework that integrates network-based prioritization, machine learning, and automated literature analysis to prioritize and rank potential therapeutic targets. Central to GETgene-AI is the G.E.T. strategy, which combines three data streams: mutational frequency (G List), differential expression (E List), and known drug targets (T List). These components are iteratively refined and ranked using the Biological Entity Expansion and Ranking Engine (BEERE), leveraging protein-protein interaction networks, functional annotations, and experimental evidence. Additionally, GETgene-AI incorporates GPT-4o, an advanced large language model, to automate literature-based ranking, reducing manual curation and increasing efficiency. In this study, we applied GETgene-AI to pancreatic cancer as a case study. The framework successfully prioritized high-priority targets such as PIK3CA and PRKCA, validated through experimental evidence and clinical relevance. Benchmarking against GEO2R and STRING demonstrated GETgene-AI's superior performance, achieving higher precision, recall, and efficiency in prioritizing actionable targets. Moreover, the framework mitigated false positives by deprioritizing genes lacking functional or clinical significance. While demonstrated on pancreatic cancer, the modular design of GETgene-AI enables scalability across diverse cancers and diseases. By integrating multi-omics datasets with advanced computational and AI-driven approaches, GETgene-AI provides a versatile and robust platform for accelerating cancer drug discovery. This framework bridges computational innovations with translational research to improve patient outcomes.

KEYWORDS

cancer, pancreatic cancer, network-based prioritization, computational biology and bioinformatics, drug target prioritization, drug target, network biology, gene prioritization

1 Introduction

Traditional chemotherapeutic agents, which non-specifically target rapidly dividing cells (Gu et al., 2023; Sun et al., 2021), are contested with the promise of targeted therapies that disrupt specific molecular pathways governing cell survival and apoptosis (Sellers and Fisher, 1999; Lim et al., 2019). Drug target discovery is pivotal for advancing cancer therapies, yet traditional approaches face three critical limitations. First, manual curation of literature and static biomedical databases struggles to scale with the complexity of modern

multi-omics data (genomic, transcriptomic, proteomic), leading to incomplete or outdated target identification (Paananen and Fortino, 2020; Zhou et al., 2022; Trajanoska et al., 2023; Lindsay, 2003; Zhou and Zhong, 2017). Second, traditional network-based prioritization, which prioritize genes based on protein-protein interaction (PPI) network centrality, oversimplify biological context by ignoring tissue-specific genomic features such as mutation frequencies and differential expression profiles (Petti et al., 2020). These limitations contribute to high failure rates in translating preclinical discoveries to clinical therapies, particularly in genetically heterogeneous cancers like pancreatic cancer. Third, reliance on single-metric approaches like fold change or mutational frequency introduces variability due to arbitrary thresholds and sample bias (McCarthy and Smyth, 2009; Dinstag and Shamir, 2020; López-Cortés et al., 2018). These gaps contribute to high failure rates in translating preclinical discoveries to clinical therapies, particularly in genetically heterogeneous cancers like pancreatic cancer (Singh et al., 2023; Sun et al., 2022; Zhu et al., 2021; Somarelli et al., 2019).

Computational advances address these challenges by integrating multi-omics data, network-based prioritization and AI-driven literature review, driving down costs, increasing precision, and expediting the development of effective therapies through *in silico* assessments (Sadybekov and Katritch, 2023; Sliwoski et al., 2014; Huan et al., 2010; Chen et al., 2006). The integration of multi-omics data contextualizes mutations within tissue-specific expression patterns, while network-based prioritization refines prioritization by mapping genes to functionally relevant pathways (Shim et al., 2015). Network-based prioritization enables researchers to analyze genomic datasets and identify critical regulatory genes implicated in cancer development (Chang et al., 2021; Sonehara and Okada, 2021). These methods prioritize disease-related genes by integrating data from PPI networks and known gene-drug associations (Mohsen et al., 2021; Zhang et al., 2021). Furthermore, network-based prioritization approaches provide the ability to efficiently process genomic information and derive meaningful insights is pivotal for identifying and visualizing relevant drug targets (Chen et al., 2013; Chen et al., 2009; Huan et al., 2010; Shim et al., 2015; Huang et al., 2012).

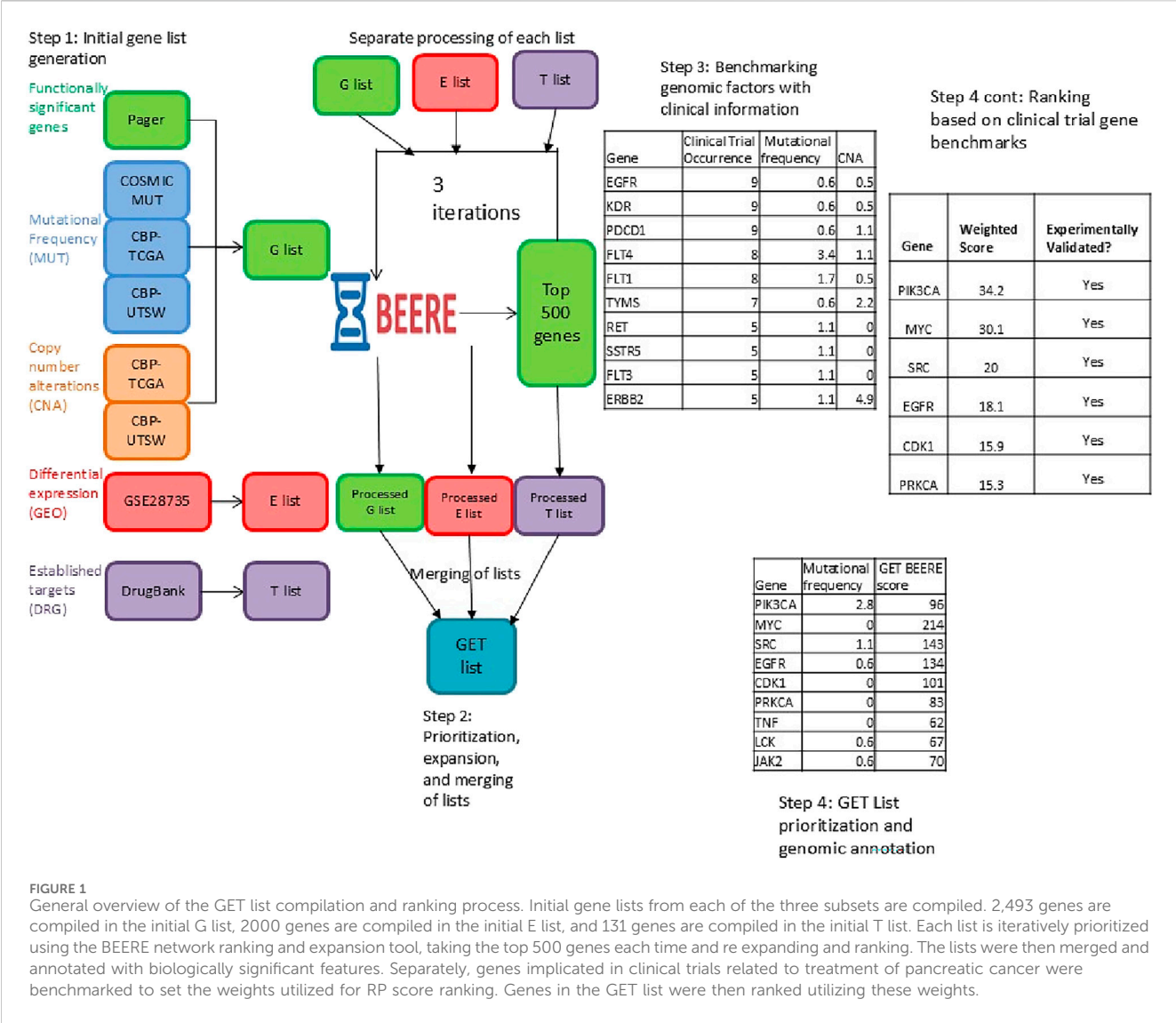
Differential gene expression is a critical method for identifying genes significantly altered between conditions, such as cancerous versus normal tissues (Bai et al., 2013; Van de Sande et al., 2023). A common approach involves calculating “fold change,” which quantifies the ratio of gene expression levels between these states (Love et al., 2014; Mutch et al., 2002). GEO2R, a tool to determine differentially expressed genes, utilizes fold change to rank genes under experimental conditions (ie. tumor versus healthy tissue comparisons) (Barrett et al., 2013). However, the arbitrary selection of fold change thresholds can introduce variability into prioritization, compromising the reliability of target identification (McCarthy and Smyth, 2009). Separately, frequency-based prioritization methods focus on genes with elevated mutational rates in disease contexts, hypothesizing these as common therapeutic targets (Dinstag and Shamir, 2020; López-Cortés et al., 2018). Frequency-based prioritization methods for gene prioritization can be prone to bias, especially due to sample selection, which can skew results (Lazzeroni et al., 2014). To address these limitations, network centrality-based prioritization has emerged as a

complementary strategy. This approach leverages gene connectivity within biological networks, offering a holistic framework for target selection by expanding gene lists and strengthening disease association metrics (Janyasupab et al., 2021; Magger et al., 2012).

Concurrently, AI-driven literature review (e.g., GPT-4) automates the synthesis of preclinical and clinical evidence, identifying targets with mechanistic and translational relevance (Liu et al., 2021; Oniani et al., 2024; Sallam, 2023; Tripathi et al., 2024). By combining these approaches, biases inherent to single-metric or fragmented datasets can be mitigated, yielding prioritized targets with mechanistic, functional, and translational relevance. (Somarelli et al., 2019; Zhu et al., 2021; Sadybekov and Katritch, 2023). LLMs can predict essential information about gene targets, including structural domains of proteins, protein structure, toxicity and adverse effects, functional significance, clinical and preclinical relevance, and treatment efficacy (Sallam, 2023; Tripathi et al., 2024). Furthermore, GPT-4 has demonstrated the ability to rival human performance in conducting literature reviews, thus streamlining the drug target prioritization process (Khraisha et al., 2024; Li et al., 2010).

In this study, we hypothesize that the utilization of network-based analysis, artificial intelligence, and biologically significant data will enable systemic prioritization of actionable therapeutic targets. Thus, we propose GETgene-AI, a framework which annotates network-based analysis with LLM enabled literature review, and biologically significant data. Central to GETgene-AI is the G.E.T. strategy, which integrates three key data streams: the G List (genes with genetic mutations, variations functionally implicated in genotype-to-phenotype association studies of the disease), the E List (disease target tissue-specific expressions of the candidate gene), and the T List (established drug targets based on reports from literature, patents, clinical trials, or existing approved drugs). Initial gene candidates are derived from heterogeneous biological datasets, including fold change, copy number alterations, and mutational frequency metrics. To mitigate biases inherent to fragmented or incomplete data, GETgene-AI employs a multi-dataset integration approach. The framework iteratively refines candidate lists through the network-based tool BEERE, which annotates and prioritizes genes with network-based centrality methods to create a high-quality, prioritized gene list. This iterative process expands and ranks candidates by evaluating their biological relevance, network centrality, and concordance with genomic aberrations, thereby improving target identification accuracy. GPT-4o is integrated into the process to improve literature review efficiency and further annotate the target list, enhancing the overall workflow. By combining traditional and *in silico* methods, GETgene-AI bridges gaps in drug discovery and facilitates the development of personalized cancer therapies.

The novel drug targets prioritized through our case study in pancreatic cancer not only offer insights into the unique molecular mechanisms driving this aggressive cancer but also present promising avenues for therapeutic intervention. While pancreatic cancer serves as a case study in this paper, the underlying methodology is adaptable to a wide range of cancers and diseases, thereby accelerating the discovery of therapeutic options.



2 Methods

In Figure 1, we show a general overview of the GETgene-AI framework.

The initial gene list is generated by employing a three-tiered strategy—comprising the Gene list (G list), Expression list (E list), and Target list (T list)—to integrate biological context into gene prioritization. The G list identifies genes with high mutational frequency, functional significance (e.g., pathway enrichment via the Kyoto Encyclopedia of Genes and Genomes (KEGG)), and genotype-phenotype associations. The E list focuses on genes exhibiting significant differential expression in pancreatic ductal adenocarcinoma (PDAC) compared to normal tissues, while the T list incorporates genes annotated as drug targets in clinical trials, patents, or approved therapies. To construct these lists, disease-specific genomic data were aggregated from public databases (e.g., TCGA, COSMIC, PAGER) and processed using GRIPPs (Gong and Chen, 2023), an iterative network-based approach that applies modality-specific thresholds to ensure robust inclusion criteria.

Following the initial gene list generation, the second step involves prioritizing and expanding these lists using the BEERE network-ranking tool. BEERE was selected for its demonstrated efficacy in filtering low-confidence data and enhancing prioritization accuracy (Yue et al., 2019), ensuring comprehensive and reliable gene sets.

A benchmark set of genes implicated in pancreatic cancer clinical trials (i.e., genes appearing as targets or biomarkers in registered interventional studies) was analyzed to evaluate which genomic and network features are most characteristic of clinically successful drug targets. This benchmark set is distinct from the T list, which consists only of genes targeted by FDA-approved drugs already indicated for pancreatic cancer. Genomic features considered included differential expression, mutation frequency, and copy number alterations, while network-based features included the BEERE scores of Gene, Expression, and Target lists. The benchmarking analysis did not alter the composition or scoring of the T list but instead provided interpretive context by identifying which factors were enriched among clinically validated targets. This analysis was further supplemented by a GPT-4-enabled literature

review, which added biological and clinical insights to the interpretation of results.

Finally, the GETgene-AI ranking is generated by integrating BEERE network rankings, annotated gene information, and insights derived from GPT-4. This multi-layered approach ensures a robust and contextually informed prioritization of potential drug targets.

Using PDAC as a case study—selected due to its poor prognosis and limited therapeutic options (Hu et al., 2021)—our framework produced quantitative data and novel insights into potential therapeutic targets, demonstrating its utility in advancing precision oncology.

2.1 Initial gene list generation

2.1.1 Compiling the gene list from genetic mutations

For the “GENE” component of our “GET” framework, we compiled three gene subsets: PAGER-NC, COSMIC-MUT, and CBP-CNA-MUT. The initial “GENE” list was compiled from the PAGER (Huang et al., 2012; Yue et al., 2018; 2022), cBioPortal (de Bruijn et al., 2023), and COSMIC (Tate et al., 2019) databases. To address potential sample biases and data incompleteness (e.g., studies failing to detect specific genes), we incorporated multiple datasets from these repositories when available. Genes associated with the term “Pancreatic Cancer” were manually curated from these databases. Empirical cutoffs were applied to prioritize genes with relevance to pancreatic cancer.

To integrate biological pathway context into gene prioritization, we utilized PAGER (Chowbina et al., 2009), which quantifies functional significance through pathway-based metrics. From PAGER, 844 candidate genes were selected heuristically using an nCoCo score threshold between 5 and 100. The nCoCo score, which measures gene set coherence by integrating co-citation and pathway data, with higher scores indicating stronger biological cohesion was constrained with a minimum of 5 (minimal coherence) and maximum of 100 (ubiquitous processes) (Huang et al., 2012; Yue et al., 2018; Yue et al., 2022).

For the cBioPortal and COSMIC databases, thresholds were defined by identifying points where mutational frequency no longer demonstrated cancer-specific significance in prior studies. From cBioPortal, 1,000 genes were selected using cutoffs of 8.2% for copy number alterations (CNA) and 2.8% for mutational frequency. The threshold for copy number alterations is significantly higher due to only 21 sets of copy number signatures being represented in 97% of tumor samples on The Cancer Genome Atlas (Steele et al., 2022). The 2.8% cutoff for mutational frequency is due to the fact that a limited amount of genes were found to be mutated in more than 5% of tumors (Sinkala, 2023). Most biologically relevant genes were found to be mutated at frequencies between 2%–20% (Lawrence et al., 2014). From COSMIC, 649 genes were compiled using a 20% mutational frequency cutoff according to the previously mentioned frequency range. Finally, candidate genes from PAGER, cBioPortal, and COSMIC were aggregated to form the “G list”, comprising 2,493 genes in total.

Sensitivity analysis was performed by testing lower and higher cutoffs for both CNA and mutational frequency. For CNA, a lower threshold of 7.3% and a higher threshold of 9.2% were applied, while for mutational frequency, thresholds of 2.2% (lower) and 3.4%

(higher) were used. For the COSMIC cancer database, a lower cutoff of 15% and a higher cutoff of 25% were applied. Genes within the top 250 of GETgene-AI were manually examined to identify those included or functionally related to genes falling within the lower and higher thresholds. The lower threshold did not identify any genes beyond those already present in the G list, whereas the higher threshold excluded the following genes: P3H2, P4HTM, PLOD3, PLOD2, P4HA1, PLOD1, PAM, PSMB5, C1QC, C1QA, and C1QB. All of these genes rank outside the top 150.

2.1.2 Compiling candidate genes for the “expression” subset

Candidate genes were prioritized by analyzing the GEO dataset GSE29735, titled “Pancreatic ductal adenocarcinoma tumor and adjacent non-tumor tissue” (Zhang et al., 2012; Zhang et al., 2013), using the GEO2R tool. Samples were categorized into tumor and non-tumor groups via the “Define groups” feature, with the tumor group defined as “human pancreatic tumor tissue patient samples” and the non-tumor group as “human pancreatic non-tumor patient samples”. The dataset comprised of 90 patient samples, evenly distributed between 45 tumor and 45 non-tumor samples. Differentially gene expression analysis was performed using GEO2R’s “analyze” function. The top 2,504 genes exhibiting log₂fc values over 0.25 were compiled into an initial “E list”. A cutoff of 0.25 was determined based on the “FindAllMarker” function provided by the R package Seurat (Wang et al., 2024). The list was subsequently processed iteratively using the BEERE software in accordance with the GRIPPs method.

2.1.3 Compiling candidate genes for the “Target” subset

Incorporating pharmacology data with network-based prioritization is a well established approach (Huang et al., 2015; Huang et al., 2012b). Building on this methodology, a set of 131 genes were identified using DrugBank (Wishart et al., 2018), a comprehensive drug and drug-target database. To extract relevant genes, the database was queried using the search terms “Pancreatic Cancer,” “Pancreatic Ductal Adenocarcinoma,” and “Neuroendocrine Pancreatic Cancer” within its drug repository. Drugs explicitly indicated for Pancreatic Cancer treatment were identified by reviewing their associated metadata, including summaries, background descriptions, indications, clinical trial references, and listed “Associated Conditions.” Each drug’s mechanism of action, therapeutic summary, and clinical trial references were manually evaluated to distinguish agents directly treating pancreatic cancer from those used for supportive care (e.g., chemotherapy relief, pain management, or sedation). For all drugs meeting the inclusion criteria, gene targets listed under their respective “Targets” section in DrugBank were compiled, resulting in 131 unique genes associated with pancreatic cancer therapeutics.

2.2 Prioritization and expansion of GET lists

To improve the specificity and biological relevance of our candidate gene lists, we implemented an iterative refinement process using the BEERE tool for prioritization and network-based expansion. The BEERE tool employs an initial ranking

algorithm and two iterative ranking algorithms—PageRank and an ant-colony algorithm—both of which have demonstrated success across diverse knowledge domains (Yue et al., 2019). Although both ranking algorithms use an iterative ranking process, they differ in how node importance weights are calculated. The PageRank algorithm assigns node importance directly from neighboring nodes. In the ant-colony algorithm nodes lose score when disseminating information and gain score upon receiving it. BEERE expands the gene list using the nearest-neighbor network constructed from protein-protein interactions in the HAPPI 2.0 database (Chen et al., 2009; 2017; Wu et al., 2012).

This workflow addresses the inherent limitations of single-dimensional analyses (e.g., relying solely on mutation or expression data) by integrating complementary biological evidence. Building on the GRIPPS framework (Gong and Chen, 2023), we developed a customized pipeline to systematically prioritize genes from three distinct categories: the combined GET list (genes ranked by aggregated mutational frequency, differential expression, and known drug-target status), the GT list (genes co-occurring in mutation and drug-target databases to highlight functionally relevant drivers), and a prioritized Expression (E) list (genes ranked exclusively by differential expression in pancreatic cancer).

The GET, GT, and E lists are expanded independently to preserve modality-specific signal during the BEERE prioritization phase. Combining them before expansion would dilute distinct biological features (e.g., mutation-specific drivers in G vs. expression-based biomarkers in E) and bias the expansion toward categories with larger initial representation, potentially overshadowing rare but high-impact genes. For example, MYC and TNF, identified through differential expression and drug-target overlap but not mutational frequency, would have been deprioritized if lists were merged prior to expansion. This systematic, modality-preserving approach enhanced the identification of potential therapeutic targets by ensuring that candidates from each evidence stream were equally represented in the final prioritization.

Each list underwent the same refinement workflow to balance comprehensiveness with specificity. First, BEERE expanded the initial gene sets by incorporating proximal interactors from protein-protein interaction (PPI) networks in the HAPPI 2.0 database, thereby capturing functionally related genes beyond those directly identified in our initial screens. Next, BEERE's network propagation and statistical ranking algorithms prioritized genes based on their network centrality and significance scores. To prevent overexpansion and maintain focus on high-confidence candidates, we empirically filtered each list to retain the top 500 genes after each prioritization cycle. This iterative process was repeated three times, as preliminary testing revealed that additional iterations caused excessive convergence of the lists, reducing their distinct biological relevance. Three iterations optimally preserved the unique profiles of each list while still enabling meaningful integration.

The independently expanded GET, GT, and E lists (each refined through three iterations of BEERE network expansion) were consolidated into an Initial GET List, which then underwent a final BEERE-based prioritization to generate the Final GET List. For comparative analysis, we also retained the previously defined

Expression List (top differentially expressed genes) and the GT List (prioritized genes from mutation–drug target overlaps). These lists were not re-derived here but carried forward for side-by-side evaluation. This tiered approach ensured that our final candidate pool retained both mechanistic diversity (genes linked to distinct biological processes) and clinical relevance (genes with actionable potential as drug targets).

The refinement process was critical to address three key challenges: (1) mitigating the high false-positive rate inherent to mutational and expression screens in heterogeneous cancers like pancreatic adenocarcinoma, (2) reconciling discrepancies between genes prioritized by individual data types (e.g., highly mutated genes often lack expression changes, and *vice versa*), and (3) ensuring functional coherence by embedding candidates within PPI networks reflective of disease biology. By iteratively refining lists through network propagation and multi-evidence integration, we enhanced the biological plausibility of candidates while preserving distinct mechanistic hypotheses for downstream validation.

2.3 GPT-4o aided literature assessment

Recent research has demonstrated that GPT-4o performs “human-like” literature reviews, particularly in screening and analyzing scientific literature (Khraisha et al., 2024). For this study, abstracts related to pancreatic cancer genes and treatments were downloaded using PubMed's “save” feature. A total of 5,091 abstracts were collected and uploaded for analysis by GPT-4o through a custom GPTo interface. Due to the data processing limitations of GPT-4o, abstracts were filtered to include only meta-analyses, clinical trials, and systematic reviews on PubMed to ensure high-quality input data.

The custom GPTo model was configured with specific instructions to rank genes based on a scoring system with a maximum score of 400 points, distributed across four categories: functional significance in pancreatic cancer, research popularity, treatment effectiveness when targeting or inhibiting the gene, and protein structure. Each category was allocated 100 points, and the resulting metric was termed the GPT-4 score. To mitigate GPT-4o's known issue of “hallucination” or the generation of inaccurate or nonexistent information, the model was explicitly instructed to base its rankings solely on the uploaded research database. Additionally, the model was required to cite articles referenced during the ranking process and provide explanations for the scores assigned to each gene in every category. GPT-4 outputs were manually verified against curated datasets to ensure biological relevance and mitigate hallucinations. Citations provided by GPT-4 were cross-referenced with PubMed to confirm validity. All cited articles were manually verified, and any errors or hallucinations were addressed by instructing the model to re-search the uploaded literature database for accurate mentions of the gene. Analyses involving database-derived information was performed on static datasets downloaded, ensuring that any subsequent database changes would not affect our reported results. Where possible, we provide accession numbers and dataset DOIs. This approach guarantees that the gene rankings and annotations presented here can be reproduced independently of future GPT-4 updates or changes to online resources.

2.4 Incorporation of clinically implicated genes and annotation of genes with factors relevant to drug target prioritization

Clinical trials are critical for evaluating the efficacy of therapeutic agents targeting specific genes. To assess the clinical relevance of prioritized genes, we quantified clinical trial activity by compiling the frequency of trials associated with each gene. Genes targeted by drugs investigated in pancreatic cancer treatment trials systemically identified through the following process: A search for the term “pancreatic cancer” was conducted on [Clinicaltrials.gov](https://clinicaltrials.gov), and all drugs listed in active or completed interventional trials for pancreatic cancer were extracted. Corresponding target genes for these drugs were then identified using DrugBank’s “Targets” section, which provides genes targeted by the drug for pancreatic cancer treatment. This process yielded 357 drugs targeting 253 unique genes. These genes were annotated with BEERE scores derived from the previously described GET lists. To enhance biological validity, the analysis integrated quantitative genomic datasets. Mutation frequency data was obtained from cBioPortal ([de Bruijn et al., 2023](#)), while protein expression profiles across tissues relevant to therapeutic safety (e.g., brain, gastrointestinal tract, liver, and kidney) were sourced from the ProteinAtlas ([Uhlén et al., 2015](#)).

Following the prioritization of the GET list and identification of clinically trialed genes, we annotated these genes with functional genomic data. Mutational frequency—a key determinant in gene ontology ranking ([Timar and Kashofer, 2020](#))—and Copy Number Alterations (CNA), a critical marker of genomic instability ([Beroukhim et al., 2010](#)), were evaluated. Mutation and CNA data were sourced from cBioPortal ([de Bruijn et al., 2023](#)) using two cohorts: the “Pancreatic Cancer (UTSW, Nat Commun 2015)” and “Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)” studies, both of which employed whole-exome sequencing for all samples. Network-based metric was also added through BEERE scores, namely the G-list score, GT-list score, E-list score, GET-list score, and the T-list score. The G, E, and T list scores are the BEERE prioritization scores derived from network-based expansion of the lists prioritized in step 2 of the methods. The GET list score is similarly from the merged GET-list detailed in step 2 of the methods. The GT-list score is a combination of the prioritized G and T scores, which aims to bring genes of higher mutational frequency into the network of the T list.

Tissue-specific expression is a vital factor in gene prioritization ([Beroukhim et al., 2010](#)). Genes with high expression in essential tissues—such as the heart, liver, gastrointestinal system, brain, and kidneys—pose a higher risk of adverse effects when targeted, necessitating their de-prioritization. Annotation of tissue expression was performed using the “RNA expression score” provided by ProteinAtlas ([Uhlén et al., 2015](#)), a comprehensive database mapping protein expression in various organs. This RNA expression score, manually calculated, measures the RNA expression levels of genes across different tissues.

2.5 GETgene-AI ranking

To unify these criteria, we developed a weighted RP score that integrates mutation frequency, copy number alterations (CNA), tissue expression, GET list scores (BEERE prioritization scores derived from

network-based expansion), E list scores, GT list scores, and clinical trial activity. Clinical trial popularity was quantified as the number of registered interventional trials testing drugs targeting a given gene for cancer therapy. Modality weights were calibrated by Spearman rank correlation between each modality-specific ranking and two independent benchmarks of therapeutic relevance: (i) the number of associated clinical trials and (ii) the frequency of reported adverse events. The benchmark set used for this analysis consisted of genes implicated in pancreatic cancer clinical trials, independent of the GET and GT lists. Correlations with clinical trial count were used to assess genomic and network features (e.g., mutation frequency, CNA frequency, GET BEERE scores), while correlations with adverse event frequency were used to assess tissue expression features (e.g., expression in brain, liver, lung, and digestive system). Modalities showing stronger monotonic associations contributed proportionally more to the final RP score, while weaker associations retained smaller weights to preserve the potential for novel candidate discovery. [Table 1](#) summarizes the relative weights of each factor in the RP score, ranked in descending order of contribution.

2.6 Mitigation of bias and false positives

To address potential sample biases and data incompleteness—such as studies failing to detect specific genes—multiple datasets from the same databases were utilized wherever possible. This redundancy ensured a more comprehensive analysis and minimized the impact of dataset-specific variability. For example, multiple studies within cBioPortal, such as “Pancreatic Cancer (UTSW, Nat Commun 2015)” and “Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas),” were analyzed concurrently to increase the reliability of mutational frequency and CNA data.

Bias from literature frequency was mitigated by not using citation counts, publication frequency, or other literature-derived popularity metrics as a direct modality in the RP score. Instead, GETgene-AI rankings are based on cancer-type-specific genomic, transcriptomic, and drug-target evidence (mutation frequency, CNA, expression, and network centrality). While genes such as PIK3CA, EGFR, PRKCA, and TNF are indeed well known, their high ranks in our framework derive from pancreatic cancer-specific data rather than their prevalence in the broader cancer literature.

Sensitivity analysis was performed by testing lower and higher cutoffs for both CNA and Mutational Frequency. A lower threshold of 7.3% and a higher threshold of 9.2% was utilized for CNA, while a lower cutoff of 2.2% and a higher cutoff of 3.4% was utilized for mutational frequency. A lower cutoff of 15% and a higher cutoff of 25% was utilized for COSMIC cancer database. Manually searching for genes within the top 250 of GETgene-ai that were included or had functionally related genes within the lower and higher thresholds. A lower threshold did not yield any genes previously not found in the G list, while the higher threshold found P3H2, P4HTM, PLOD3, PLOD2, P4HA1, PLOD1, PAM, PSMB5, C1QC, C1QA, C1QB, to be genes excluded due to higher thresholding. These genes all rank outside of the top 150.

To further enhance the accuracy of the prioritization process, each gene within the top 250 ranked by RP score was manually verified through a literature review to confirm its role in cancer biology. This step was critical in identifying and eliminating false

TABLE 1 Weights each modality was assigned for calculation of the RP score in GETgene-AI.

Modality of ranking	Weighted score
GT list score	0.329
CNA(CBIOPORTAL UTSW NAT COMMUN 2015)	0.201
Expression list score	0.088
GET list score	0.085
Mutation frequency (cBioporta ITCGA PanCancerAtlas)	0.079
CNA(CBIOPORTAL TCGA PANCANCERATLAS)	0.048
Mutation frequency (Cbiportal UTSW Nat Commun 2015)	−0.023
Brain expression score	−0.054
Kidney expression score	−0.081
Gastrointestinal expression score	−0.095
Liver expression score	−0.101

positives. Notably, no genes within the top 250 were found to be false positives, validating the robustness of the RP scoring methodology.

Additionally, hallucination errors from GPT-4o were mitigated through a structured training approach. The model was instructed to explicitly cite a source used in the calculation of each gene's ranking score. These citations were manually evaluated for accuracy and relevance, ensuring that the ranking process was grounded in verifiable scientific evidence. This dual-layered validation—automated scoring combined with manual review—was integral to maintaining the integrity and reliability of the gene prioritization framework.

2.7 Statistical methods

Spearman correlation coefficients were computed to assess the alignment of GPT-4o rankings with network-derived rankings. The Spearman correlation between the GPT-4 score and the Weighted Score was 0.291, indicating some significance. Interestingly, GPT-4 score is more strongly correlated with all BEERE list ranking scores, with 0.478 between GPT-4 score and Expression list score, 0.457 between GPT-4 score and Combined weighted score of all BEERE lists, 0.454 correlations between GPT-4 score and GET list score, and 0.444 between GPT-4 score and GT list score. These results indicate that the GPT-4 score is more similar to that of standard network prioritization techniques, which may be a result of the training data utilized.

2.8 Comparing research relevance to rank on GETgene-AI

To compare the popularity to the rankings of each gene in both the GPT-4 Score and the RP scores, the amount of results contained on PubMed when searching “Gene name Pancreatic Cancer” were compiled and used for the GPT-LIT score, and the RP-LIT score.

The GPT-LIT score is the GPT4-score divided by the amount of publications on PubMed, while the RP-lit score is the RP-score divided by the amount of publications on PubMed. Genes with no functional relationship to cancer in any way were excluded from the rankings to remove false positives.

3 Results

3.1 GETgene-AI rankings and validations

We observe the highest ranked genes according to GETgene-AI in [Table 2](#).

During the iterative ranking process, genes lacking functional relevance to cancer were systematically deprioritized. For instance, genes that ranked highly due to algorithmic artifacts but lacked experimental validation or literature support were ranked lower than genes with experimental validation or literature support. The final candidate set was defined as the top 250 genes ranked by RP score. This threshold was selected to enable manual literature verification for each gene, ensuring that all final candidates could be cross-checked for pancreatic cancer-specific evidence and therapeutic relevance. Expanding the list beyond this size would have substantially increased the manual verification burden without proportionally improving the quality of candidates for downstream analysis. This approach allowed us to maintain both methodological rigor and practical feasibility while focusing on the most highly ranked genes.

PIK3CA emerged as the highest-ranked gene on our list. It encodes the enzyme PI3K, which regulates critical cellular processes such as growth, metabolism, proliferation, and apoptosis ([Conway et al., 2019](#)). *PIK3CA* also modulates downstream effectors, including *AKT* and *mTOR* ([Ala, 2022](#)), and preclinical studies demonstrate that mutations in this gene sensitize cancers to dual *PI3K/mTOR* inhibitors ([Zhang et al., 2021](#)), underscoring its therapeutic potential. Notably, *PIK3CA*-null tumors exhibit heightened susceptibility to T-cell surveillance *in vitro* ([Sivaram et al., 2019](#)), while its inhibition in pancreatic cancer models initiates tumorigenesis ([Payne et al., 2015](#)), highlighting its dual role in progression and therapy.

MYC, the second highest-ranked gene, achieved its position due to its top GET list score, reflecting its network centrality among the 500 most expressed, clinically relevant, and frequently mutated genes. Overexpression of *c-MYC* is a hallmark of aggressive pancreatic cancer, where it binds promoter regions of oncogenic targets ([Hayashi et al., 2021](#)). Despite its pivotal regulatory role, *MYC*'s complex protein structure poses therapeutic challenges, resulting in a lower GT list score. Recent advances in small-molecule inhibitors, however, show preclinical promise.

SRC ranks as the third-highest gene on our list, driven by its high scores in both the GET list and Expression list modalities. Inhibition of *SRC* in pancreatic cancer has been shown to reverse chemoresistance to pyroptosis in both *in vitro* and *in vivo* studies ([Su et al., 2023](#)). Aberrant *SRC* activity promotes tumorigenesis and is frequently associated with poor prognosis in pancreatic ductal adenocarcinoma (PDAC) ([Poh and Ernst, 2023](#)). Several *SRC*-targeting therapies are currently under clinical investigation ([Hilbig, 2008](#)).

TABLE 2 Highest 20 genes ranked on GETGENE-AI. Weighted score is RP score, CHAT GPT score is GPT4 score.

Gene	RP score	CHATGPT score	GT list score	Mutation frequency (cBioportal TCGA PanCancer Atlas)	RP-LIT score	GPT-LIT score	GET list score	Expression list score
PIK3CA	34.8	310	58.7	2.8	0.199	1.771	96	97
MYC	30.1	330	9.5	0.0	0.032	0.349	214	210
SRC	20.0	320	0.0	1.1	0.044	0.711	143	144
EGFR	18.2	320	2.4	0.6	0.010	0.171	134	133
CDK1	15.9	305	15.3	65.4	0.134	2.563	30	7
PRKCA	15.3	305	3.0	0.0	1.702	25.556	101	102
TNF	12.1	270	2.4	0.0	0.013	0.292	83	86
LCK	11.5	220	1.7	0.0	1.274	24.444	62	60
JAK2	10.6	285	1.0	0.6	0.082	2.192	67	67
MAPK1	10.3	305	11.6	3.4	0.139	4.122	7	7
AURKB	9.1	295	0.0	0.6	0.008	0.246	70	70
KRAS	8.7	220	1.7	1.7	0.335	8.462	48	47
MAPK8	7.8	295	0.0	0.0	0.002	0.068	121	117
MTOR	7.1	220	1.7	0.0	0.588	18.333	52	52
ITGA4	6.9	220	4.3	0.6	2.298	73.333	40	37
TOP2A	6.9	310	10.2	1.1	0.215	9.688	0	0
CHEK1	6.7	220	1.7	0.0	0.128	4.231	46	45
BCL2	6.2	220	1.7	0.6	0.012	0.418	41	41
PRKCB	6.0	250	1.4	0.6	1.004	41.667	60	58
ERBB4	5.5	220	3.4	0.6	0.184	7.333	81	83

EGFR is the fourth highest-ranked gene, attributed to its high GET list and Expression list scores. *EGFR* is also implicated in tumorigenesis, particularly in lung and breast cancer (Sigismund et al., 2018). Anti-*EGFR* agents have shown significant clinical promise, despite associated adverse effects (Verma et al., 2020).

KRAS ranks 12th on our list, despite its prominence in pancreatic cancer research, with over 4,545 PubMed articles on *KRAS* mutations in pancreatic cancer. Its lower ranking is primarily due to a low expression score. The *KRAS* oncogene plays a critical role in the initiation and maintenance of pancreatic tumors (Luo, 2021). *KRAS* mutations are present in over 90% of PDAC cases, but therapeutic inhibition remains highly challenging, with effective inhibitors only recently being discovered (Bannoura et al., 2021).

CDK1 ranks fifth on our list, largely due to its high scores in both the GET and Expression lists. *CDK1* is strongly correlated with prognosis and is highly expressed in pancreatic cancer tissue, as well as in response to gemcitabine, an approved pancreatic cancer drug (Xu et al., 2023). Additionally, inhibition of *CDK1*, along with *CDK2* and *CDK5*, has been shown to overcome IFN- γ -triggered acquired resistance in pancreatic tumor immunity (Huang et al., 2021).

PRKCA ranks seventh on our list. It encodes protein kinase C and is mutated in various cancers. *PRKCA*'s high ranking is attributed to its strong GET and Expression list scores, as well as its extremely low organ expression score. It is strongly associated

with the activation of the protein translation initiation pathway (Rosenberg et al., 2018) and is a hallmark mutation in chordoid gliomas (Jiang et al., 2019). *PRKCA* also contributes to susceptibility to pancreatic cancer through the peroxisome proliferator-activated receptor (*PPAR*) signaling pathway, which plays a key role in pancreatic cancer development and progression (Liu et al., 2020). Inhibition of *PRKCA* has demonstrated antitumor activity in patients with advanced non-small cell lung cancer (NSCLC) (Villalona-Calero et al., 2004).

TNF is the eighth highest-ranked gene on our list. Tumor Necrosis Factor (*TNF*) upregulation is associated with invasion and immunomodulation in pancreatic cancer (Wiedmann et al., 2023). *TNF*-mutated macrophages have also been shown to promote aggressive cancer behaviors through lineage reprogramming (Tu et al., 2021).

LCK ranks ninth on our list. This gene is expressed in tumor cells and plays a key role in T-cell development (Bommhardt et al., 2019). High *LCK* protein expression has been associated with improved patient survival in cancer (Cancer Genome Atlas Network, 2015). Despite its biological relevance, *LCK* has only four PubMed publications discussing its role in pancreatic cancer as of May 2024. Its identification as a high-priority target demonstrates GETgene-AI's ability to prioritize genes with strong biological relevance but limited literature prominence.

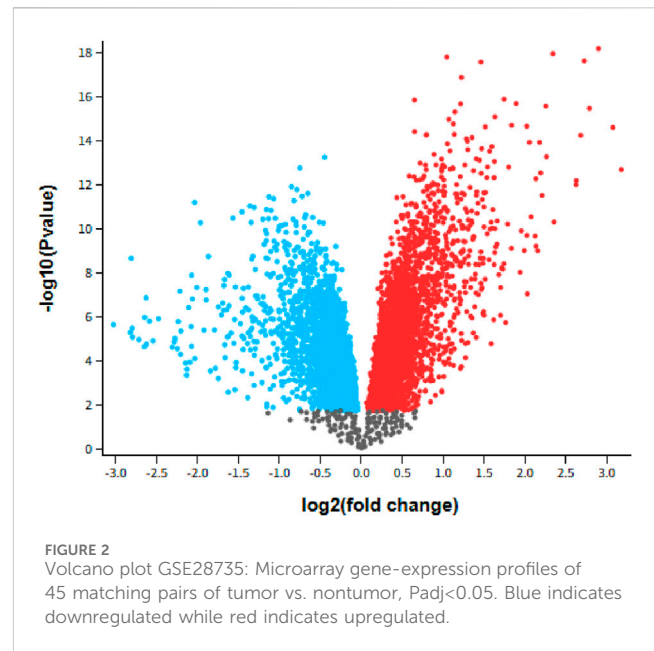
ITGA4 ranks 15th on our list. It has an extremely low organ expression score and only four PubMed articles discussing its role in pancreatic cancer. *ITGA4* has potential as an independent prognostic indicator for patient survival and has been linked to the *PI3K/AKT* pathway (Faleiro et al., 2021). Its identification as a high-priority target further highlights GETgene-AI's capability to prioritize genes with strong biological relevance despite limited literature attention.

KCNA ranks 34th on our list. Notably, there are no PubMed publications describing its relation to pancreatic cancer, and only three publications mention its role in cancer in general. The identification of *KCNA* as a high-priority target underscores GETgene-AI's ability to prioritize genes with strong biological relevance but minimal literature prominence. *KCNA* exhibits differentially high expression in stomach and lung cancers and is positively correlated with infiltrated immune cells and survival rates (Angi et al., 2023).

3.2 Comparing GETgene-AI to other frameworks

We benchmarked GETgene-AI against two other frameworks: one focused on differential expression analysis and the other on network-based gene prioritization. For the differential expression comparison, we selected GEO2R, utilizing the GSE28735 dataset, which was integrated into the 'Expression list' component of our GET lists. Genes were ranked based on their log-fold change (log-fc), representing the difference in gene expression between tumor and non-tumor groups. In the GEO2R list, the top-ranked genes were *PNLIPRP1* and *PNLIPRP2*, both of which encode pancreatic lipase-related proteins critical for digestion and fat absorption (Zhu et al., 2021). However, these genes are not considered viable targets for pancreatic cancer. The third-ranked gene, *IAPP* (Islet Amyloid Polypeptide), has been shown to lack tumor suppressor functionality, and loss of *IAPP* signaling is not associated with pancreatic cancer (Taylor et al., 2023). Among the top 50 genes identified by GEO2R, 30 were experimentally validated as relevant to pancreatic cancer. In contrast, GETgene-AI prioritized 49 experimentally validated targets within its top 50, representing a 38% improvement over GEO2R. GEO2R's limitations, including the absence of mutational frequency analysis, functional impact assessment, network-based analysis, and adverse effect evaluation, hinder its utility in drug target discovery. In comparison, GETgene-AI leverages statistical filtering and incorporates genomic information, significantly enhancing both the efficiency and quality of gene prioritization. Figure 2 presents a volcano plot illustrating the log₂ (fold change) distributions for the analyzed genes.

For the network-based comparison, we employed STRING, a database that integrates protein-protein interaction data (Szkarczyk et al., 2023), focusing specifically on the KEGG pathway hsa0512 (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2025). Genes were ranked based on node degree, a measure of the number of interactions a protein has within the network (Bozhilova et al., 2019). The highest-ranked gene in the STRING list was *AKT1*, a protein kinase known to stimulate cell growth and proliferation (Grassilli et al., 2020). However, *AKT1* has been shown to resist



inhibition by shifting its metabolic activity from glycolysis to mitochondrial respiration (Arasanz et al., 2019). Additionally, it exhibits a low mutational frequency of only 1% in a cohort of 19,784 patients with various tumors (Millis et al., 2016). Due to its low mutational frequency and the challenges associated with its inhibition, *AKT1* was ranked 33rd by GETgene-AI. Among the top 50 genes prioritized by STRING, 46 were experimentally validated for relevance to pancreatic cancer, whereas GETgene-AI identified 49 experimentally validated genes within its top 50, demonstrating a 6% improvement over STRING. STRING's limitations, such as its inability to account for mutational frequency and other critical factors in drug target identification, result in a narrower focus, with only 81 targets prioritized compared to the more comprehensive analysis provided by GETgene-AI. Figure 3 illustrates the network constructed using STRING.

Comparing GETgene-AI to GEO2R and STRING, our framework demonstrates a 38% improvement over GEO2R and a 6% improvement over STRING in the rate of experimental validation of the top 50 genes on each list. In Figure 4, we observe the differences in the percentage of experimentally validated targets out of the top 50.

GETgene-AI was also compared to OpenTarget, an integrative AI-based prioritization platform (Koscielny et al., 2017). We compared GETgene-AI's rankings to those generated by OpenTargets for pancreatic cancer, focusing on the top 15 genes from each tool. While there was overlap in high-confidence drivers (e.g., *KRAS*, *TP53*, *SMAD4*, *BRCA2*), several key differences emerged that highlight the value of GETgene-AI's multi-modal integration.

OpenTargets ranked genes such as *POLE* and *POLD1* highly despite their low mutation frequency in pancreatic cancer datasets (*POLE* absent in one TCGA cohort; *POLD1* <1% in UTSW CNA and mutation frequency). GETgene-AI deprioritized these genes due to the lack of mutational enrichment and limited pancreatic-specific evidence, avoiding inflation from literature-based or pathway-only associations.

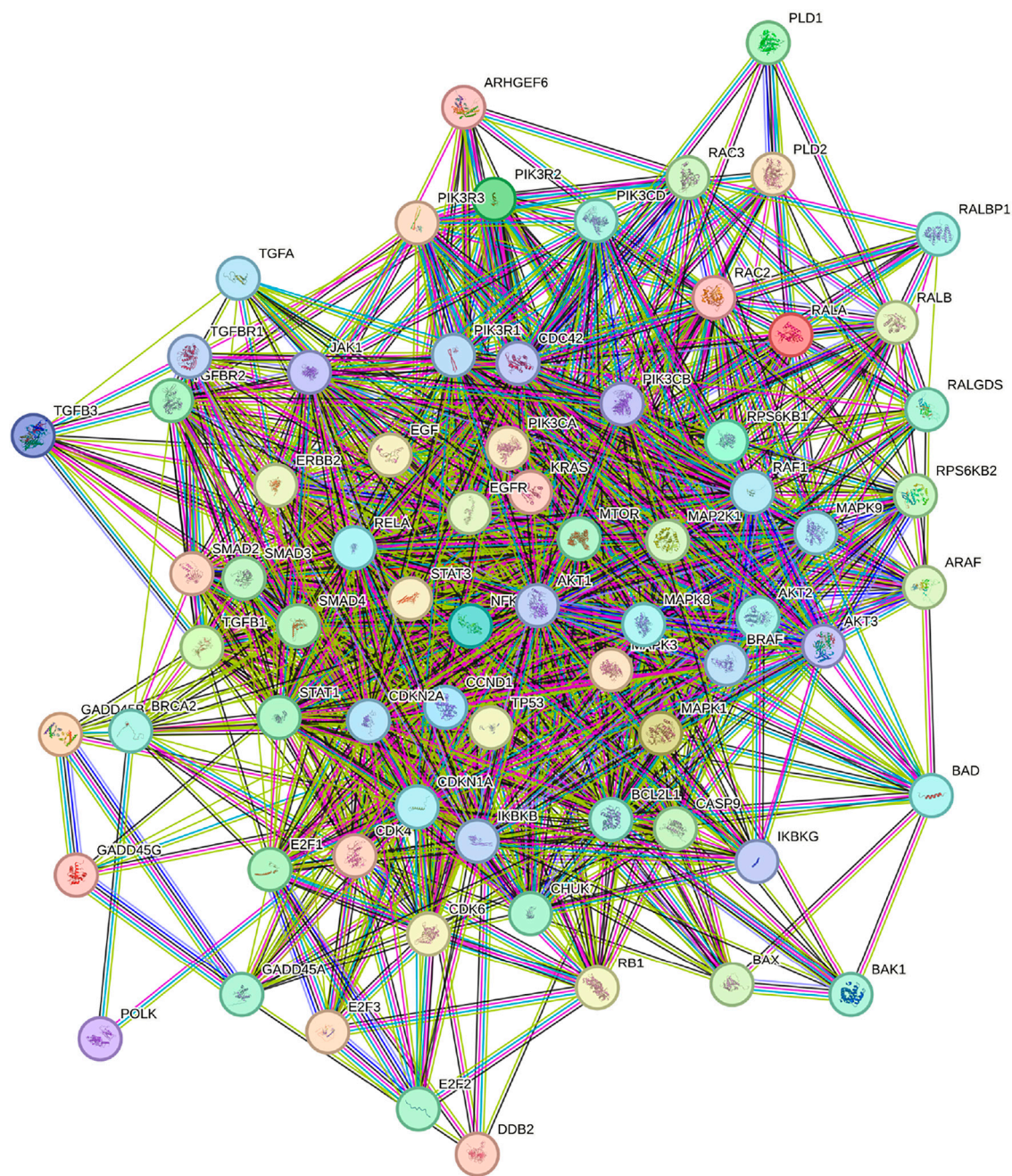


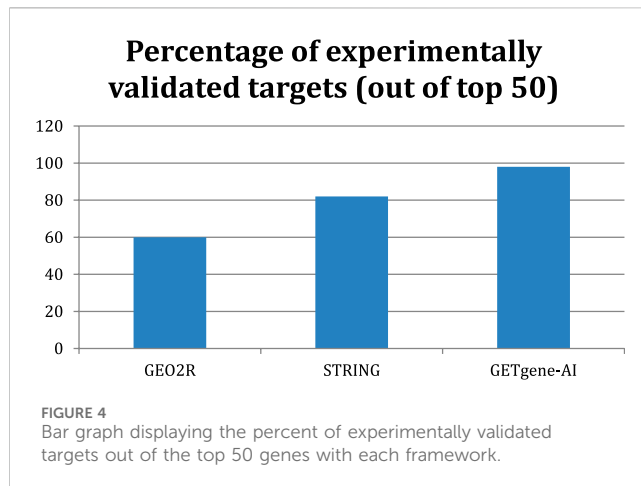
FIGURE 3 Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3days structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate predicted Protein-protein interactions. Light green edges represent text mining, black represents co-expression, and light purple represents protein homology.

Conversely, GETgene-AI prioritized genes such as MYC, SRC, EGFR, and CDK1, which have strong differential expression and drug-target relevance in pancreatic cancer but were absent from OpenTargets’ top list.

These differences indicate that OpenTargets may overweight generalized associations, whereas GETgene-AI incorporates cancer-type-specific genomic, transcriptomic, and therapeutic data, leading

to rankings more aligned with the biological and clinical context of pancreatic cancer.

In Table 3, we observe the ranking overlap for the top 15 genes for all three frameworks. The top 15 highest ranked targets in both GETgene-AI and STRING have all been experimentally validated within pancreatic cancer, but 8 of the highest ranking targets in the GEO2R approach have not.



3.3 Enhancement provided by AI

GPT-4o was utilized to conduct a comprehensive literature assessment for our gene list. Although its output was not incorporated into the final weighted score, the GPT-4o scores demonstrated strong correlations with both the weighted score and all three GET list scores. Notably, GPT-4o prioritized genes such as *MYC* and *SRC*, reflecting their well-documented prominence in the scientific literature. This complemented GETgene-AI's approach, which relies on network mutational analysis for gene prioritization. To minimize the inclusion of false positives in the GPT-4o scoring process, we instructed GPT-4o to directly cite articles from its internal database. While GPT-4o did not exhibit a higher rate of experimental validation compared to manual methods, it

significantly reduced the time required for literature review by 80%. All cited articles were subsequently manually verified to ensure accuracy.

The RP-LIT score and GPT-4o score showed a high degree of correlation, with extremely similar rankings for each gene. Based on Spearman correlation analysis, the GPT-4o score (out of 400) exhibited a correlation coefficient of +0.457 with the weighted score, indicating a statistically significant relationship. Table 4 provides a detailed comparison of the ranking differences between the GPT-4o score and the GET ranking score, highlighting the alignment and discrepancies between the two approaches.

3.4 False positives and limitations

False positives are an inherent risk in large-scale computational analyses. The GETgene-AI framework addresses this challenge through iterative refinement and the systematic exclusion of genes lacking functional or experimental support. Future validation efforts will focus on further refining these rankings through targeted experimental studies. Additionally, the literature assessment provided by generative AI is expected to improve as AI technology advances and our model is trained on more experimental data, thereby minimizing inaccuracies or "hallucinations" in the generated outputs.

To mitigate false positives, genes without functional relevance to cancer were systematically excluded. For instance, genes that ranked highly due to algorithmic artifacts but lacked experimental validation or literature support were deprioritized. Examples include *ITGA4* and *PRKCB*, both of which have fewer than 10 PubMed articles discussing their role in pancreatic cancer. These genes were ranked lower than many well-established targets due to their low scores in the GET, GT, and Expression

TABLE 3 Top 15 genes from GETGENE-AI, STRING, and GEO2R and their status as experimentally validated drug targets.

GETGENE-AI top genes	Experimentally validated?	STRING top genes	Experimentally validated?	GEO2R top genes	Experimentally validated?
PIK3CA	Yes	AKT1	Yes	PNLIPRP1	No
MYC	Yes	TP53	Yes	PNLIPRP2	No
SRC	Yes	KRAS	Yes	IAPP	No
EGFR	Yes	PTEN	Yes	CTRC	No
CDK1	Yes	SRC	Yes	GP2	Yes
PRKCA	Yes	STAT3	Yes	CEL	No
TNF	Yes	EGFR	Yes	CPA2	Yes
LCK	Yes	MTOR	Yes	ALB	Yes
JAK2	Yes	BCL2	Yes	CUZD1	Yes
MAPK1	Yes	PIK3CA	Yes	ERP27	No
MTOR	Yes	CDKN2A	Yes	CLPS	Yes
AURKB	Yes	HRAS	Yes	SERPINI2	Yes
KRAS	Yes	CCND1	Yes	PLA2G1B	Yes
MAPK8	Yes	NFKB1	Yes	CELA2A	No
TOP2A	Yes	CDKN1A	Yes	CELA2B	No

TABLE 4 Top 20 highest ranked genes based off of GPT4 score compared to their ranks in GET and their status as experimentally validated drug targets.

Gene	GPT4-score ranking	GET ranking	Experimental validation?	Citation
MYC	1	2	Yes	Zhang et al. (2024)
SRC	2	3	Yes	Su et al. (2023)
EGFR	3	4	Yes	Wu et al. (2023)
TERT	4	27	Yes	Campa et al. (2015)
RRM2	5	21	Yes	Li et al. (2022)
PIK3CA	6	1	Yes	Payne et al. (2015)
TOP2A	7	16	Yes	Pei et al. (2018)
NTRK1	8	22	Yes	Cheng et al. (2013)
PTGS2	9	25	Yes	Hingorani et al. (2003)
EGF	10	30	Yes	Sheng et al. (2020)
CDK1	11	5	Yes	Huang et al. (2021)
MAPK1	12	10	Yes	Si et al. (2023)
KRAS	13	13	Yes	Timar and Kashofer (2020)
MTOR	14	11	Yes	Stanciu et al. (2022)
MSLN	15	37	Yes	Hu et al. (2024)
RET	16	28	Yes	Bhamidipati et al. (2023)
AKT1	17	31	Yes	Arasanz et al. (2019)
JAK2	18	9	Yes	Huang et al. (2022)
MET	19	34	Yes	Pothula et al. (2020)
PDCD1	20	38	Yes	Marabelle et al. (2020)

lists, which prioritize targets with robust experimental or literature support during the RP score calculation process.

This study has several limitations. First, the top-ranked targets identified by GETgene-AI require further experimental validation, which is a critical next step to confirm their biological and therapeutic relevance. Second, the reliance on publicly available datasets may introduce biases due to incomplete or inconsistent annotations. These limitations highlight the need for further experimental validation and the incorporation of more comprehensive datasets to enhance the accuracy and reliability of the framework.

3.5 Broader implications and generalizability

While the current study focuses on pancreatic cancer, the GETgene-AI framework can be readily adapted to other cancers or diseases with access to similar genomic and clinical data resources. Future studies will explore its application to breast and lung cancers by employing the same systematic process described in this work. The GETgene-AI framework integrates literature review, large-scale sequencing data, and network centrality scores, providing a comprehensive approach to drug target prioritization. Additionally, its reliance on computational methods for prioritization and the elimination of statistically insignificant data ensures that the framework is both scalable and efficient, making it suitable for broader applications in biomedical research.

4 Discussion

Through the application of GETgene-AI to pancreatic cancer, we have identified several promising drug targets, including PIK3CA, PRKCA, LCK, MAPK8, ITGA4, PRKCB, and KCNA1, warranting further investigation. These targets display strong pancreatic cancer-specific genomic and transcriptomic evidence, high network centrality in PPI analyses, and have not been extensively reported in the pancreatic cancer literature despite their biological relevance in our analysis.

GETgene-AI’s approach to drug target prioritization integrates literature review, large-scale sequencing data, network-based centrality scoring, and assessment of potential adverse effects through organ expression scores. This multifaceted implementation offers a scalable and comprehensive framework for drug target prioritization, which can be readily adapted to other cancers with similar data availability. Furthermore, GETgene-AI’s ability to systematically deprioritize genes with low mutational relevance underscores its superiority in efficiently narrowing down actionable and biologically relevant targets. Slight variations of cutoffs utilized for the compilation and prioritization of the GET lists did not result in significant variations of the final rankings or scores of the final GETgene-AI gene list.

In contrast to recent methods that rely largely on AI-driven network analysis alone (e.g., an AI-Driven Network Biology pipeline

TABLE 5 genes highlighted in the discussion section labeled by significance.

Gene	Significance
PIK3CA	Investigated in PDAC clinical trials
PRKCA	Investigated in PDAC preclinical models (<i>in vitro</i> or <i>in vivo</i>)
LCK	Investigated in PDAC preclinical models (<i>in vitro</i> or <i>in vivo</i>)
MAPK8	Investigated in PDAC preclinical models (<i>in vitro</i> or <i>in vivo</i>)
ITGA4	Novel and unstudied in PDAC
PRKCB	Novel and unstudied in PDAC
KCNA1	Novel and unstudied in PDAC

identifying SRC as a therapeutic target in pancreatic cancer) (Zhang and Chen, 2025), GETgene-AI offers a more automated and modular framework. Our approach not only evaluates protein–protein interaction networks but also incorporates tissue-specific gene expression and mutation frequency analyses, and integrates these modalities through distinct G, E, and T lists before merging. This enables multi-dimensional prioritization grounded in genomic, transcriptomic, and therapeutic evidence. In future extensions, the modular nature of GETgene-AI allows easy incorporation of additional evaluation modules—such as differential tissue analysis, motif-based mutation enrichment, or epigenetic regulation scores—each processed independently in their own list and then integrated via our weighted RP score. This design ensures adaptability and enables seamless expansion of the framework to accommodate new modalities as the data landscape evolves.

4.1 Contributions and limitations provided by GPT4o

GPT-4o significantly enhanced the efficiency of literature-based ranking by automating the review and prioritization of scientific abstracts. This approach increased the efficiency of literature review by over 80%. However, inherent challenges, such as the risk of hallucination, necessitated manual verification to ensure the accuracy of the results. While GPT-4o provides substantial value, its integration into research workflows should be approached cautiously, with safeguards implemented to mitigate potential errors. Additionally, training GPT-4o on more experimental data in the future will further improve its accuracy and reliability in prioritization tasks.

4.2 Future directions

While the current study focuses on cancer applications, future research will expand the scope of the GETgene-AI framework. We plan to validate its utility in additional cancer types, such as breast and lung cancer, and explore its applicability to non-cancerous disease contexts, including neurodegenerative disorders like Alzheimer’s and Parkinson’s. By integrating computational methods with large-scale genomic data, the GETgene-AI framework addresses critical gaps in drug discovery, accelerating the identification of actionable targets and advancing the development of personalized therapeutic strategies.

Future work will prioritize experimental validation of top-ranked targets, such as *PIK3CA* and *PRKCA*, using CRISPR-mediated knockouts in pancreatic cancer cell lines. Subsequent *in vitro* drug response assays will evaluate the therapeutic potential of these targets. Additionally, we aim to refine the framework by incorporating multi-omics datasets (e.g., proteomics, metabolomics) and enhancing its ability to predict adverse effects through improved organ expression profiling. ce of these targets.

5 Conclusion

The GET framework represents a significant advancement in computational drug discovery, integrating network-based prioritization with machine learning to prioritize actionable therapeutic targets efficiently. Genes highlighted through our case study in pancreatic cancer such as *PRKCA*, *LCK*, *ITGA4*, and *PRKCB* are novel targets that require further exploration. While this study focuses on pancreatic cancer, the GETGENE-AI framework is adaptable to other cancers and diseases, offering a modular and versatile approach for target discovery. GPT4o enhanced the efficiency and accuracy of literature-based ranking, reducing manual workload and aligning well with network-based rankings. However, its reliance on manual verification underscores the need for cautious integration into automated pipelines. By refining target discovery methods, the GETGENE-AI framework paves the way for personalized therapeutic strategies and accelerates the translational research in oncology. Future work will focus on expanding the framework to other cancers, improving ranking metrics, and integrating multi-omics datasets to enhance its predictive power. Future iterations of GETgene-AI aim to integrate multi-omics datasets, such as single-cell RNA-seq and metabolomics, to capture greater biological complexity. Table 5 indicates the significance of each gene labeled as novel.

Data availability statement

The data presented in the study are deposited in the <https://github.com/alphamind-club/GETGENE-AI> repository.

Author contributions

AG: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing. JC: Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. JYC thanks the generous support of the startup fund to the Systems Pharmacology AI

Research Center at UAB and NIH common fund grant award U54-OD036472, which partially supported this research.

Acknowledgments

Both authors thank the administrative support of AlphaMind Club for making this mentored research possible. The authors acknowledge the use of ChatGPT in improving the structure and readability of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

References

- Ala, M. (2022). Target c-Myc to treat pancreatic cancer. *Cancer Biol. Ther.* 23, 34–50. doi:10.1080/15384047.2021.2017223
- Angi, B., Muccioli, S., Szabó, I., and Leanza, L. (2023). A meta-analysis study to infer voltage-gated K⁺ Channels prognostic value in different cancer types. *Antioxid. Basel Switz.* 12, 573. doi:10.3390/antiox12030573
- Arasanz, H., Zuazo, M., Santamaría, E., Bocanegra, A. I., Gato-Cañas, M., Fernández-Hinojal, G., et al. (2019). Adaption of pancreatic cancer cells to AKT1 inhibition induces the acquisition of cancer stem-cell like phenotype through upregulation of mitochondrial functions. *Ann. Oncol.* 30, v11. doi:10.1093/annonc/mdz238.036
- Bai, J. P. F., Alekseyenko, A. V., Statnikov, A., Wang, I.-M., and Wong, P. H. (2013). Strategic applications of gene expression: from drug discovery/development to bedside. *AAPS J.* 15, 427–437. doi:10.1208/s12248-012-9447-1
- Bannoura, S. F., Uddin, Md. H., Nagasaka, M., Fazili, F., Al-Hallak, M. N., Philip, P. A., et al. (2021). Targeting KRAS in pancreatic cancer: new drugs on the horizon. *Cancer Metastasis Rev.* 40, 819–835. doi:10.1007/s10555-021-09990-2
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Beroukhir, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. doi:10.1038/nature08822
- Bhamidipati, D., Yedururi, S., Huse, J., Chinapuvvula, S. V., Wu, J., and Subbiah, V. (2023). Exceptional responses to seliparitinib in RET fusion-driven metastatic pancreatic cancer. *JCO Precis. Oncol.* 7, e2300252. doi:10.1200/PO.23.00252
- Bommhardt, U., Schraven, B., and Simeoni, L. (2019). Beyond TCR signaling: emerging functions of lck in cancer and immunotherapy. *Int. J. Mol. Sci.* 20, 3500. doi:10.3390/ijms20143500
- Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., and Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinforma.* 20, 446. doi:10.1186/s12859-019-3036-6
- Campa, D., Rizzato, C., Stolzenberg-Solomon, R., Pacetti, P., Vodicka, P., Cleary, S. P., et al. (2015). TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *Int. J. Cancer* 137, 2175–2183. doi:10.1002/ijc.29590
- Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. *Cell* 161, 1681–1696. doi:10.1016/j.cell.2015.05.044
- Chang, L., Ruiz, P., Ito, T., and Sellers, W. R. (2021). Targeting pan-essential genes in cancer: challenges and opportunities. *Cancer Cell* 39, 466–479. doi:10.1016/j.ccell.2020.12.008
- Chen, J. Y., Shen, C., and Sivachenko, A. Y. (2006). Mining Alzheimer disease relevant proteins from integrated protein interactome data. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 367–378.
- Chen, J. Y., Mamidipalli, S., and Huan, T. (2009). HAPPI: an online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics* 10, S16. doi:10.1186/1471-2164-10-S1-S16
- Chen, J. Y., Piquette-Miller, M., and Smith, B. P. (2013). Network medicine: finding the links to personalized therapy. *Clin. Pharmacol. Ther.* 94, 613–616. doi:10.1038/clpt.2013.195
- Chen, J. Y., Pandey, R., and Nguyen, T. M. (2017). HAPPI-2: a comprehensive and high-quality map of human annotated and predicted protein interactions. *BMC Genomics* 18, 182. doi:10.1186/s12864-017-3512-1
- Cheng, Y., Diao, D., Zhang, H., Song, Y.-C., and Dang, C.-X. (2013). Proliferation enhanced by NGF-NTRK1 signaling makes pancreatic cancer cells more sensitive to 2DG-induced apoptosis. *Int. J. Med. Sci.* 10, 634–640. doi:10.7150/ijms.5547
- Chowbina, S. R., Wu, X., Zhang, F., Li, P. M., Pandey, R., Kasamsetty, H. N., et al. (2009). HPD: an online integrated human pathway database enabling systems biology studies. *BMC Bioinforma.* 10, S5. doi:10.1186/1471-2105-10-S11-S5
- Conway, J. R., Herrmann, D., Evans, T. J., Morton, J. P., and Timpson, P. (2019). Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine. *Gut* 68, 742–758. doi:10.1136/gutjnl-2018-316822
- de Bruijn, I., Kundra, R., Mastrogiacono, B., Tran, T. N., Sikina, L., Mazor, T., et al. (2023). Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal. *Cancer Res.* 83, 3861–3867. doi:10.1158/0008-5472.CAN-23-0816
- Dinstag, G., and Shamir, R. (2020). PRODIGY: personalized prioritization of driver genes. *Bioinformatics* 36, 1831–1839. doi:10.1093/bioinformatics/btz815
- Faleiro, I., Roberto, V. P., Demirkol Canli, S., Fraunhofer, N. A., Iovanna, J., Gure, A. O., et al. (2021). DNA methylation of PI3K/AKT pathway-related genes predicts outcome in patients with pancreatic cancer: a comprehensive bioinformatics-based study. *Cancers* 13, 6354. doi:10.3390/cancers13246354
- Gong, E., and Chen, J. Y. (2023). Prioritizing complex disease genes from heterogeneous public databases. doi:10.1101/2023.02.09.527562
- Grassilli, S., Brugnoli, F., Lattanzio, R., Buglioni, S., and Bertagnolo, V. (2020). Vav1 down-modulates Akt2 expression in cells from pancreatic ductal adenocarcinoma: nuclear Vav1 as a potential regulator of akt related malignancy in pancreatic cancer. *Biomedicines* 8, 379. doi:10.3390/biomedicines8100379
- Gu, L., Hickey, R. J., and Malkas, L. H. (2023). Therapeutic targeting of DNA replication stress in cancer. *Genes* 14, 1346. doi:10.3390/genes14071346
- Hayashi, A., Hong, J., and Iacobuzio-Donahue, C. A. (2021). The pancreatic cancer genome revisited. *Nat. Rev. Gastroenterol. Hepatol.* 18, 469–481. doi:10.1038/s41575-021-00463-z
- Hilbig, A. (2008). “Src kinase and pancreatic cancer.” in *Pancreatic cancer* (Berlin, Heidelberg: Springer Berlin Heidelberg), 179–185. doi:10.1007/978-3-540-71279-4_19
- Hingorani, S. R., Petricoin, E. F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M. A., et al. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* 4, 437–450. doi:10.1016/s1535-6108(03)00309-x
- Hu, J.-X., Zhao, C.-F., Chen, W.-B., Liu, Q.-C., Li, Q.-W., Lin, Y.-Y., et al. (2021). Pancreatic cancer: a review of epidemiology, trend, and risk factors. *World J. Gastroenterol.* 27, 4298–4321. doi:10.3748/wjg.v27.i27.4298

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI utilized as part of the drug target analysis approach. Generative AI was also utilized to revise manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Hu, J., Wang, J., Guo, X., Fan, Q., Li, X., Li, K., et al. (2024). MSLN induced EMT, cancer stem cell traits and chemotherapy resistance of pancreatic cancer cells. *Heliyon* 10, e29210. doi:10.1016/j.heliyon.2024.e29210
- Huan, T., Wu, X., and Chen, J. Y. (2010). Systems biology visualization tools for drug target discovery. *Expert Opin. Drug Discov.* 5, 425–439. doi:10.1517/17460441003725102
- Huang, H., Wu, X., Pandey, R., Li, J., Zhao, G., Ibrahim, S., et al. (2012a). C²Maps: a network pharmacology database with comprehensive disease-gene-drug connectivity relationships. *BMC Genomics* 13, S17. doi:10.1186/1471-2164-13-S6-S17
- Huang, H., Wu, X., Sonachalam, M., Mandape, S. N., Pandey, R., MacDorman, K. F., et al. (2012b). PAGED: a pathway and gene-set enrichment database to enable molecular phenotype discoveries. *BMC Bioinforma.* 13, S2. doi:10.1186/1471-2105-13-S15-S2
- Huang, H., Nguyen, T., Ibrahim, S., Shantharam, S., Yue, Z., and Chen, J. Y. (2015). DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinforma.* 16, S4. doi:10.1186/1471-2105-16-S13-S4
- Huang, J., Chen, P., Liu, K., Liu, J., Zhou, B., Wu, R., et al. (2021). CDK1/2/5 inhibition overcomes IFNG-mediated adaptive immune resistance in pancreatic cancer. *Gut* 70, 890–899. doi:10.1136/gutjnl-2019-320441
- Huang, B., Lang, X., and Li, X. (2022). The role of IL-6/JAK2/STAT3 signaling pathway in cancers. *Front. Oncol.* 12, 1023177. doi:10.3389/fonc.2022.1023177
- Janyasupab, P., Suratanee, A., and Plaimas, K. (2021). Network diffusion with centrality measures to identify disease-related genes. *Math. Biosci. Eng.* 18, 2909–2929. doi:10.3934/mbe.2021147
- Jiang, H., Fu, Q., Song, X., Ge, C., Li, R., Li, Z., et al. (2019). HDGF and PRKCA upregulation is associated with a poor prognosis in patients with lung adenocarcinoma. *Oncol. Lett.* 18, 4936–4946. doi:10.3892/ol.2019.10812
- Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci. Publ. Protein Soc.* 28, 1947–1951. doi:10.1002/pro.3715
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27
- Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., and Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* 53, D672–D677. doi:10.1093/nar/gkac909
- Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., and Hadfield, K. (2024). Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. *Res. Synth. Methods* 15, 616–626. doi:10.1002/jrsm.1715
- Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparian, R., et al. (2017). Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 45, D985–D994. doi:10.1093/nar/gkw1055
- Lawrence, M. S., Stojanov, P., Mermel, C. H., Garraway, L. A., Golub, T. R., Meyerson, M., et al. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501. doi:10.1038/nature12912
- Lazzeroni, L. C., Lu, Y., and Belitskaya-Lévy, I. (2014). P-values in genomics: apparent precision masks high uncertainty. *Mol. Psychiatry* 19, 1336–1340. doi:10.1038/mp.2013.184
- Li, J., Zhu, X., and Chen, J. Y. (2010). Discovering breast cancer drug candidates from biomedical literature. *Int. J. Data Min. Bioinforma.* 4, 241–255. doi:10.1504/IJDMB.2010.033519
- Li, W., Chen, Q., Gao, W., and Zeng, H. (2022). ARID1A promotes chemosensitivity to gemcitabine in pancreatic cancer through epigenetic silencing of RRM2. *Pharm* 77, 224–229. doi:10.1691/ph.2022.1881
- Lim, B., Greer, Y., Lipkowitz, S., and Takebe, N. (2019). Novel apoptosis-inducing agents for the treatment of cancer, a new arsenal in the toolbox. *Cancers* 11, 1087. doi:10.3390/cancers11081087
- Lindsay, M. A. (2003). Target discovery. *Nat. Rev. Drug Discov.* 2, 831–838. doi:10.1038/nrd1202
- Liu, X., Qian, D., Liu, H., Abbruzzese, J. L., Luo, S., Walsh, K. M., et al. (2020). Genetic variants of the peroxisome proliferator-activated receptor (PPAR) signaling pathway genes and risk of pancreatic cancer. *Mol. Carcinog.* 59, 930–939. doi:10.1002/mc.23208
- Liu, Z., Roberts, R. A., Lal-Nag, M., Chen, X., Huang, R., and Tong, W. (2021). AI-based language models powering drug discovery and development. *Drug Discov. Today* 26, 2593–2607. doi:10.1016/j.drudis.2021.06.009
- López-Cortés, A., Paz-y-Miño, C., Cabrera-Andrade, A., Barigye, S. J., Munteanu, C. R., González-Díaz, H., et al. (2018). Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. *Sci. Rep.* 8, 16679. doi:10.1038/s41598-018-35149-1
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Luo, J. (2021). KRAS mutation in pancreatic cancer. *Semin. Oncol.* 48, 10–18. doi:10.1053/j.seminoncol.2021.02.003
- Magger, O., Waldman, Y. Y., Rupp, E., and Sharan, R. (2012). Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol.* 8, e1002690. doi:10.1371/journal.pcbi.1002690
- Marabelle, A., Le, D. T., Ascierto, P. A., Di Giacomo, A. M., De Jesus-Acosta, A., Delord, J.-P., et al. (2020). Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 38, 1–10. doi:10.1200/JCO.19.02105
- McCarthy, D. J., and Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Oxf. Engl.* 25, 765–771. doi:10.1093/bioinformatics/btp053
- Millis, S. Z., Ikeda, S., Reddy, S., Gatalica, Z., and Kurzrock, R. (2016). Landscape of phosphatidylinositol-3-kinase pathway alterations across 19 784 diverse solid tumors. *JAMA Oncol.* 2, 1565–1573. doi:10.1001/jamaoncol.2016.0891
- Mohsen, H., Gunasekharan, V., Qing, T., Seay, M., Surovtseva, Y., Negahban, S., et al. (2021). Network propagation-based prioritization of long tail genes in 17 cancer types. *Genome Biol.* 22, 287. doi:10.1186/s13059-021-02504-x
- Mutch, D. M., Berger, A., Mansourian, R., Rytz, A., and Roberts, M.-A. (2002). The limit fold change model: a practical approach for selecting differentially expressed genes from microarray data. *BMC Bioinforma.* 3, 17. doi:10.1186/1471-2105-3-17
- Oniani, D., Hilsman, J., Zang, C., Wang, J., Cai, L., Zawala, J., et al. (2024). Emerging opportunities of using large language models for translation between drug molecules and indications. *Sci. Rep.* 14, 10738. doi:10.1038/s41598-024-61124-0
- Paananen, J., and Fortino, V. (2020). An omics perspective on drug target discovery platforms. *Brief. Bioinform.* 21, 1937–1953. doi:10.1093/bib/bbz122
- Payne, S. N., Maher, M. E., Tran, N. H., Van De Hey, D. R., Foley, T. M., Yueh, A. E., et al. (2015). PIK3CA mutations can initiate pancreatic tumorigenesis and are targetable with PI3K inhibitors. *Oncogenesis* 4, e169. doi:10.1038/oncsis.2015.28
- Pei, Y.-F., Yin, X.-M., and Liu, X.-Q. (2018). TOP2A induces malignant character of pancreatic cancer through activating β -catenin signaling pathway. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864, 197–207. doi:10.1016/j.bbdis.2017.10.019
- Petti, M., Bizzarri, D., Verrienti, A., Falcone, R., and Farina, L. (2020). Connectivity significance for disease gene prioritization in an expanding universe. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 2155–2161. doi:10.1109/TCBB.2019.2938512
- Poh, A. R., and Ernst, M. (2023). Functional roles of SRC signaling in pancreatic cancer: recent insights provide novel therapeutic opportunities. *Oncogene* 42, 1786–1801. doi:10.1038/s41388-023-02701-x
- Pothula, S. P., Xu, Z., Goldstein, D., Pirola, R. C., Wilson, J. S., and Apte, M. V. (2020). Targeting HGF/c-MET Axis in pancreatic cancer. *Int. J. Mol. Sci.* 21, 9170. doi:10.3390/ijms21239170
- Rosenberg, S., Simeonova, I., Bielle, F., Verreault, M., Bance, B., Le Roux, I., et al. (2018). A recurrent point mutation in PRKCA is a hallmark of chordoid gliomas. *Nat. Commun.* 9, 2371. doi:10.1038/s41467-018-04622-w
- Sadybekov, A. V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685. doi:10.1038/s41586-023-05905-z
- Sallam, M. (2023). ChatGPT utility in healthcare Education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthc. Basel Switz.* 11, 887. doi:10.3390/healthcare11060887
- Sellers, W. R., and Fisher, D. E. (1999). Apoptosis and cancer drug targeting. *J. Clin. Invest.* 104, 1655–1661. doi:10.1172/JCI9053
- Sheng, W., Shi, X., Lin, Y., Tang, J., Jia, C., Cao, R., et al. (2020). Musashi2 promotes EGF-induced EMT in pancreatic cancer via ZEB1-ERK/MAPK signaling. *J. Exp. Clin. Cancer Res. CR* 39, 16. doi:10.1186/s13046-020-1521-4
- Shim, J. E., Hwang, S., and Lee, I. (2015). Pathway-dependent effectiveness of network algorithms for gene prioritization. *PLOS ONE* 10, e0130589. doi:10.1371/journal.pone.0130589
- Si, H., Zhang, N., Shi, C., Luo, Z., and Hou, S. (2023). Tumor-suppressive miR-29c binds to MAPK1 inhibiting the ERK/MAPK pathway in pancreatic cancer. *Clin. Transl. Oncol.* 25, 803–816. doi:10.1007/s12094-022-02991-9
- Sigmund, S., Avanzato, D., and Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Mol. Oncol.* 12, 3–20. doi:10.1002/1878-0261.12155
- Singh, N., Vayer, P., Tanwar, S., Poyet, J.-L., Tsaioun, K., and Villoutreix, B. O. (2023). Drug discovery and development: introduction to the general public and patient groups. *Front. Drug Discov.* 3, 1201419. doi:10.3389/fddsv.2023.1201419
- Sinkala, M. (2023). Mutational landscape of cancer-driver genes across human cancers. *Sci. Rep.* 13, 12742. doi:10.1038/s41598-023-39608-2
- Sivaram, N., McLaughlin, P. A., Han, H. V., Petrenko, O., Jiang, Y.-P., Ballou, L. M., et al. (2019). Tumor-intrinsic PIK3CA represses tumor immunogenicity in a model of pancreatic cancer. *J. Clin. Invest.* 129, 3264–3276. doi:10.1172/JCI123540
- Slivoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi:10.1124/pr.112.007336
- Somarelli, J. A., Boddy, A. M., Gardner, H. L., DeWitt, S. B., Tuohy, J., Megquier, K., et al. (2019). Improving cancer drug discovery by studying cancer across the tree of life. *Mol. Biol. Evol.* 37, 11–17. doi:10.1093/molbev/msz254
- Sonehara, K., and Okada, Y. (2021). Genomics-driven drug discovery based on disease-susceptibility genes. *Inflamm. Regen.* 41, 8. doi:10.1186/s41232-021-00158-7

- Stanciu, S., Ionita-Radu, F., Stefani, C., Miricescu, D., Stanescu-Spinu, I.-I., Greabu, M., et al. (2022). Targeting PI3K/AKT/mTOR signaling pathway in pancreatic cancer: from molecular to clinical aspects. *Int. J. Mol. Sci.* 23, 10132. doi:10.3390/ijms231710132
- Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., Haase, K., et al. (2022). Signatures of copy number alterations in human cancer. *Nature* 606, 984–991. doi:10.1038/s41586-022-04738-6
- Su, L., Chen, Y., Huang, C., Wu, S., Wang, X., Zhao, X., et al. (2023). Targeting Src reactivates pyroptosis to reverse chemoresistance in lung and pancreatic cancer models. *Sci. Transl. Med.* 15, eabl7895. doi:10.1126/scitranslmed.abl7895
- Sun, Y., Liu, Y., Ma, X., and Hu, H. (2021). The influence of cell cycle regulation on chemotherapy. *Int. J. Mol. Sci.* 22, 6923. doi:10.3390/ijms22136923
- Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* 12, 3049–3062. doi:10.1016/j.apsb.2022.02.002
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. doi:10.1093/nar/gkac1000
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947. doi:10.1093/nar/gky1015
- Taylor, A. J., Panzhinskiy, E., Orban, P. C., Lynn, F. C., Schaeffer, D. F., Johnson, J. D., et al. (2023). Islet amyloid polypeptide does not suppress pancreatic cancer. *Mol. Metab.* 68, 101667. doi:10.1016/j.molmet.2023.101667
- Timar, J., and Kashofer, K. (2020). Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Rev.* 39, 1029–1038. doi:10.1007/s10555-020-09915-5
- Trajanoska, K., Bhéer, C., Taliun, D., Zhou, S., Richards, J. B., and Mooser, V. (2023). From target discovery to clinical drug development with human genetics. *Nature* 620, 737–745. doi:10.1038/s41586-023-06388-8
- Tripathi, S., Gabriel, K., Tripathi, P. K., and Kim, E. (2024). Large language models reshaping molecular biology and drug development. *Chem. Biol. Drug Des.* 103, e14568. doi:10.1111/cbdd.14568
- Tu, M., Klein, L., Espinet, E., Georgomanolis, T., Wegwitz, F., Li, X., et al. (2021). TNF- α -producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. *Nat. Cancer* 2, 1185–1203. doi:10.1038/s43018-021-00258-w
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. doi:10.1126/science.1260419
- Van de Sande, B., Lee, J. S., Mutasa-Gottgens, E., Naughton, B., Bacon, W., Manning, J., et al. (2023). Applications of single-cell RNA sequencing in drug discovery and development. *Nat. Rev. Drug Discov.* 22, 496–520. doi:10.1038/s41573-023-00688-4
- Verma, H. K., Kampalli, P. K., Lakkakula, S., Chalikhonda, G., Bhaskar, L. V. K. S., and Pattnaik, S. (2020). A retrospective look at anti-EGFR agents in pancreatic cancer therapy. *Curr. Drug Metab.* 20, 958–966. doi:10.2174/1389200220666191122104955
- Villalona-Calero, M. A., Ritch, P., Figueroa, J. A., Otterson, G. A., Belt, R., Dow, E., et al. (2004). A phase I/II study of LY900003, an antisense inhibitor of protein kinase C- α , in combination with cisplatin and gemcitabine in patients with advanced non-small cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 6086–6093. doi:10.1158/1078-0432.CCR-04-0779
- Wang, W., Li, C., Dai, Y., Wu, Q., and Yu, W. (2024). Unraveling metabolic characteristics and clinical implications in gastric cancer through single-cell resolution analysis. *Front. Mol. Biosci.* 11, 1399679. doi:10.3389/fmolb.2024.1399679
- Wiedmann, L., De Angelis Rigotti, F., Vaquero-Siguero, N., Donato, E., Espinet, E., Moll, L., et al. (2023). HAPLN1 potentiates peritoneal metastasis in pancreatic cancer. *Nat. Commun.* 14, 2353. doi:10.1038/s41467-023-38064-w
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037
- Wu, X., Huang, H., Wei, T., Pandey, R., Reinhard, C., Li, S. D., et al. (2012). Network expansion and pathway enrichment analysis towards biologically significant findings from microarrays. *J. Integr. Bioinforma.* 9, 213. doi:10.2390/biecoll-jib-2012-213
- Wu, F., He, J., Deng, Q., Chen, J., Peng, M., Xiao, J., et al. (2023). Neuroglobin inhibits pancreatic cancer proliferation and metastasis by targeting the GNAI1/EGFR/AKT/ERK signaling axis. *Biochem. Biophys. Res. Commun.* 664, 108–116. doi:10.1016/j.bbrc.2023.04.080
- Xu, X., Ding, Y., Jin, J., Xu, C., Hu, W., Wu, S., et al. (2023). Post-translational modification of CDK1–STAT3 signaling by fisetin suppresses pancreatic cancer stem cell properties. *Cell Biosci.* 13, 176. doi:10.1186/s13578-023-01118-z
- Yue, Z., Zheng, Q., Neylon, M. T., Yoo, M., Shin, J., Zhao, Z., et al. (2018). PAGER 2.0: an update to the pathway, annotated-list and gene-signature electronic repository for Human Network Biology. *Nucleic Acids Res.* 46, D668–D676. doi:10.1093/nar/gkx1040
- Yue, Z., Willey, C. D., Hjelmeland, A. B., and Chen, J. Y. (2019). BEERE: a web server for biomedical entity expansion, ranking and explorations. *Nucleic Acids Res.* 47, W578–W586. doi:10.1093/nar/gkz428
- Yue, Z., Slominski, R., Bharti, S., and Chen, J. Y. (2022). PAGER web APP: an interactive, online gene set and network interpretation tool for functional genomics. *Front. Genet.* 13, 820361. doi:10.3389/fgene.2022.820361
- Zhang, A., and Chen, J. Y. (2025). AI-driven network biology identifies SRC as a therapeutic target in metastatic pancreatic adenocarcinoma. *Intell. Oncol.* 1 (3), 233–243. doi:10.1016/j.intonc.2025.06.004
- Zhang, G., Schetter, A., He, P., Funamizu, N., Gaedcke, J., Ghadimi, B. M., et al. (2012). DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 7, e31507. doi:10.1371/journal.pone.0031507
- Zhang, G., He, P., Tan, H., Budhu, A., Gaedcke, J., Ghadimi, B. M., et al. (2013). Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 19, 4983–4993. doi:10.1158/1078-0432.CCR-13-0209
- Zhang, H., Ferguson, A., Robertson, G., Jiang, M., Zhang, T., Sudlow, C., et al. (2021). Benchmarking network-based gene prioritization methods for cerebral small vessel disease. *Brief. Bioinform.* 22, bbab006. doi:10.1093/bib/bbab006
- Zhang, H., Sun, Y., Wang, Z., Huang, X., Tang, L., Jiang, K., et al. (2024). ZDHHC20-mediated S-palmitoylation of YTHDF3 stabilizes MYC mRNA to promote pancreatic cancer progression. *Nat. Commun.* 15, 4642. doi:10.1038/s41467-024-49105-3
- Zhou, S.-F., and Zhong, W.-Z. (2017). Drug design and discovery: principles and applications. *Molecules* 22, 279. doi:10.3390/molecules22020279
- Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., et al. (2022). Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* 50, D1398–D1407. doi:10.1093/nar/gkab953
- Zhu, G., Fang, Q., Zhu, F., Huang, D., and Yang, C. (2021a). Structure and function of pancreatic lipase-related protein 2 and its relationship with pathological states. *Front. Genet.* 12, 693538. doi:10.3389/fgene.2021.693538
- Zhu, L., Jiang, M., Wang, H., Sun, H., Zhu, J., Zhao, W., et al. (2021b). A narrative review of tumor heterogeneity and challenges to tumor drug therapy. *Ann. Transl. Med.* 9, 1351. doi:10.21037/atm-21-1948