GETgene-AI: A Framework for Prioritizing Actionable Cancer Drug Targets

Adrian Gu[1] and Jake Y Chen[1, 2]

1 AlphaMind Club, Birmingham, AL, [2] Systems Pharmacology AI Research Center, School of Medicine, the University of Alabama at Birmingham, Birmingham, AL

ABSTRACT

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.

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 (Lim et al., 2019; Sellers & Fisher, 1999). 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 & Fortino, 2020; Zhou et al., 2022; Trajanoska et al., 2023; Lindsay, 2003; Zhou & 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). Third, fragmented prioritization—reliance on single-metric approaches like fold change or mutational frequency—introduces variability due to arbitrary thresholds and sample bias (McCarthy & Smyth, 2009; Dinstag & 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 and expediting the development of effective therapies through in silico assessments (Sadybekov & Katritch, 2023; Sliwoski et al., 2014; Huan et al., 2010). These approaches harness the power of neural networks, expansive genomic datasets, and sophisticated machine learning algorithms to predict genes of interest with a precision unattainable by traditional methods (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). 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). Together, these components resolve biases inherent to fragmented analyses: multi-omics data anchors targets in biological context, networks prioritize functional relevance, and AI-driven literature review ensures scalability in evidence synthesis. (Somarelli et al., 2019; Zhu et al., 2021; Sadybekov & Katritch, 2023).

Multi-Omics data integrates genes based on genomic information (Eg. Differential Gene Expression and Frequency of Mutation). 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). However, the arbitrary selection of fold change thresholds can introduce variability into prioritization, potentially compromising the reliability of target identification (McCarthy & Smyth, 2009). Tools like GEO2R exemplify this approach, leveraging fold change to rank genes under experimental conditions, such as tumor versus healthy tissue comparisons (Barrett et al., 2013). Separately, frequency-based prioritization methods focus on genes with elevated mutational rates in disease contexts, hypothesizing these as common therapeutic targets (Dinstag & Shamir, 2020; López-Cortés et al., 2018). While valuable, 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).

Network-based prioritization enables researchers to analyze genomic datasets and identify critical regulatory genes implicated in cancer development (Chang et al., 2021; Sonehara & 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). The ability to efficiently process genomic information and derive meaningful insights is pivotal for identifying and visualizing relevant drug targets, underscoring the importance of network-based prioritization (Chen et al., 2013; Chen et al., 2009; Huan et al., 2010). Network-based

prioritization methods are particularly favored for gene prioritization, as they utilize established biological networks to identify genes associated with specific diseases (Shim et al., 2015; Huang et al., 2012). As noted earlier, traditional network-based prioritization methods lack tissue specificity (Petti et al., 2020), necessitating the integration of genomic data.

As noted earlier, AI-Driven Literature review through the use of Large Language Models (LLMs), such as GPT-4, have emerged as another transformative *in silico* approach to drug prioritization. (Liu et al., 2021; Oniani et al., 2024). 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 a critical component of 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.

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 & 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.

In the third step, genes implicated in clinical trials were incorporated and annotated with factors relevant to drug target prioritization. These included factors such as clinical trial popularity, differential expression, mutation frequency, and copy number alterations. This

process is further enriched by leveraging GPT-4 enabled literature review, which provides additional context and insights into the biological and clinical relevance of the identified genes.

Finally, the GETgene-AI ranking is generated by creating a weighted score called RP-score which integrates genomic information, BEERE list scores, and expression in important organs to assess the efficacy of targeting a gene. The weights for each variable were determined by calculating Spearman correlations between the specific variable and clinical trial popularity, defined as the number of cancer treatment–related clinical trials in which the gene was targeted.

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.

Initial gene list generation

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. We integrated nCoCo score since it measures coherence by integrating co-citation and pathway data, with higher scores indicating stronger biological cohesion (Huang et al., 2012; Yue et al., 2018, 2022). Groups with nCoCo scores below 5 lacked coherent pathway annotations or shared functional roles (e.g., random co-citation clusters), reducing their relevance to disease mechanisms. Groups with nCoCO scores above 100 represented ubiquitous biological processes (e.g., "cellular metabolism") with broad roles across tissues, diluting disease-specific signal.

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 signitures 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. To prevent data bias, datasets with differing mutational frequency within samples of the same genes were selected. Integration of multiple datasets was also utilized to prevent sample bias. Finally, candidate genes from PAGER, cBioPortal, and COSMIC were aggregated to form the “G list”, comprising 2,493 genes in total.

While we did not formally reprocess the entire dataset with varied cutoffs, preliminary assessments suggested that slight variations (e.g., ±5% in mutation frequency thresholds) did not significantly alter the prioritized gene rankings. This stability indicates that GETgene-AI’s prioritization is driven by biologically relevant signals rather than arbitrary thresholds.

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, 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 logfc 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.

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.

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 & Chen, 2023), we developed a customized pipeline to systemically 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) . This systematic approach ensured a robust and multifaceted analysis of candidate genes, enhancing the identification of potential therapeutic targets.

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 refined outputs were consolidated into an Initial GET List, which underwent a final BEERE-based prioritization to generate the Final GET List. For comparative analysis, we also derived two additional lists: the GT List (prioritized genes from mutation-drug target overlaps) and the Expression List (top differentially expressed genes). 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).

This 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.

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.

GPT-4o 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-4o outputs were manually verified against curated datasets to ensure biological relevance and mitigate hallucinations. Citations provided by GPT-4o 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.

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, 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 & 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.

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.

GETgene-AI ranking

To unify these criteria, we developed a weighted RP score that integrates mutation frequency, CNA, mutational frequency, GET list scores, tissue expression, E list scores, GT list scores, and clinical trial popularity (number of trials testing drugs targeting a given gene for cancer therapy). Spearman correlation analysis was performed to assess pair wise relationships between these variables. The Spearman correlation is a widely used statistical tool for prioritizing drug target discovery, as it quantifies the strength and direction of monotonic associations between variables (Tanoli et al., 2025). Unlike the Pearson correlation, which measures linear proportional relationships, Spearman’s method evaluates rank-based correlations, making it less sensitive to outliers—a critical advantage when analyzing noisy biological datasets (Rodríguez-Pérez & Bajorath, 2021). Additionally, Spearman correlation captures both linear and nonlinear monotonic relationships, enabling researchers to identify diverse patterns in genomic data that might otherwise be overlooked (Penrod & Moore, 2014).

These features collectively make it a versatile choice for exploring complex biological associations in drug target discovery pipelines. The RP score weights were derived from the directional Spearman correlation coefficients (ρ) between clinical trial popularity and each genomic/contextual variable. Unlike absolute correlations, this approach preserves the sign of the relationship, allowing variables with positive associations (e.g., higher mutation frequency → more trials) to enhance the RP score, while variables with negative associations (e.g., higher tissue expression → fewer trials) reduce it. This directional weighting reflects the observed biological and clinical trends, where certain genomic features either promote or deter therapeutic targeting. The RP score provides a multi-dimensional ranking mechanism for therapeutic target prioritization. GTlistscore received the highest weight because it integrates established drug targets and functionally or frequently mutated genes into its score. Table 1 summarizes the relative weights of each factor in the RP score, ranked in descending order of contribution.

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. We combined data sources after thresholding as a form of preprocessing to ensure stable

rankings, and reducing sensitivity to arbitrary thresholds. Mutational and CNA data were merged from multiple studies to reduce cohort specific bias.

While we did not formally reprocess the entire dataset with varied cutoffs, preliminary assessments suggested that slight variations (e.g., ±5% in mutation frequency thresholds) did not significantly alter the prioritized gene rankings. This stability indicates that GETgene-AI's prioritization is driven by biologically relevant signals rather than arbitrary thresholding.

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 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.

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.

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.

RESULTS

GETgene-AI Rankings and Validations

The highest ranked genes according to GETgene-AI in table 2.

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 & Ernst, 2023). Several *SRC*-targeting therapies are currently under clinical investigation (Hilbig, 2008).

EGFR is the fourth highest-ranked gene, attributed to its high GET list and Expression list scores. *EGFR* is not ranked higher due to having lower mutational and copy number alteration frequencies than other high ranking targets. *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 twelfth on our list, despite its prominence in pancreatic cancer research, with over 4,545 PubMed articles on *KRAS* mutations in pancreatic cancer. However, it is ranked lower due to its low differential expression, being ranked 27562 on GEO2R. Additionally, KRAS is considered undruggable due to the lack of classic drug binding sites, further increasing difficulty in targeting it (Zhu et al., 2022). 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).

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.

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. We chose to benchmark off of precision, defined as the percentage of genes found to be functionally significant in literature. PubMed was utilized to search for literature in this benchmark. 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 (Szklarczyk et al., 2023), focusing specifically on the KEGG pathway hsa0512 (Kanehisa, 2019; Kanehisa et al., 2025; Kanehisa & Goto, 2000). 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, 47 were experimentally validated for relevance to pancreatic cancer, whereas GETgene-AI identified 49 experimentally validated genes within its top 50, demonstrating an 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.

In table 3, we observe the ranking overlap for the top 50 genes for all three frameworks. 49 out of the top 50 highest ranked targets in GETgene-AI have been experimentally validated, while 47 were validated in STRING. However, only 30 of the highest ranking targets were validated in GEO2R.

Comparing GETgene-AI to GEO2R and STRING, our framework demonstrates a 38% improvement over GEO2R 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.

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%. Human oversight was essential to ensure accuracy and reliability of GPT-4o in verifying the literature and reasoning of GPT-4o before incorporation into the pipeline.

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.

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.

DISCUSSION

Through the application of GETgene-AI to pancreatic cancer, we have identified several under researched promising drug targets: PIK3CA, PRKCA, LCK, MAPK8, ITGA4, PRKCB, and KCNA1. We have validated our hypothesis that the utilization of network-based analysis, artificial intelligence, and biologically significant data will enable systemic prioritization of

actionable therapeutic targets. 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. Additionally, the integration of an assessment of potential adverse effects through organ expression scores strengthens its applicability to finding targets with low adverse events in clinical trials. 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.

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 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. Third, disease specific factors cannot be directly considered with GETgene-AI, such as protein aggregation in Alzheimer's Disease (Pluta & Ułamek-Kozioł, 2020).

Contributions and Limitations Provided by GPT4o

GPT-4o significantly enhanced the efficiency of literature-based ranking by helping automate the review and prioritization of scientific abstracts. This approach significantly increased the efficiency of literature review. However, inherent challenges, such as the risk of hallucination, necessitated manual verification to ensure the accuracy of the results. As such, artificial intelligence is utilized as a complementary tool for searching and assessing literature, rather than a replacement to human expertise (Blanco-González et al., 2023). 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.

Future Directions

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.

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.

CONCLUSIONS

The GETGENE-AI 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-omnics 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.

Data availability

All files containing genes and drugs utilized in the GETGENE-AI process have been uploaded to our Github repository https://github.com/alphamind-club/GETGENE-AI. “GEO2R assessment”, “GETGENE-AIvsGET”, “GETvsGPT4vsSTRINGvsGEO2R”, and “STRING Assessment” contain experimental validation assessments performed to compare the top 50 genes from each approach. Other files are the initial lists of genes utilized to form each of the GET lists before cutoffs were applied. ChatGPT’s GPT4o model was used for the GPT4o literature assessment score, which requires a monthly subscription. The custom GPT4o model made for this study is: Research GPT. ProteinAtlas was used to find the expression levels.

Acknowledgment

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

REFERENCES

Ala, M. (2022). Target c-Myc to treat pancreatic cancer. *Cancer Biology & Therapy*, *23*(1), 34–50. https://doi.org/10.1080/15384047.2021.2017223

Angi, B., Muccioli, S., Szabò, I., & Leanza, L. (2023). A Meta-Analysis Study to Infer Voltage-Gated K+ Channels Prognostic Value in Different Cancer Types. *Antioxidants (Basel, Switzerland)*, *12*(3), 573. https://doi.org/10.3390/antiox12030573

Arasanz, H., Zuazo, M., Santamaría, E., Bocanegra, A. I., Gato-Cañas, M., Fernández-Hinojal, G., Hernández-Saez, C., Fernández-Irigoyen, J., Santamaria, E., Rodríguez-Ortigosa, C., Kochan, G., Vera, R., & Escors, D. (2019). Adaption of pancreatic cancer cells to AKT1 inhibition induces the acquisition of cancer stem-cell like phenotype through upregulation of mitochondrial functions. *Annals of Oncology*, *30*, v11. https://doi.org/10.1093/annonc/mdz238.036

Bai, J. P. F., Alekseyenko, A. V., Statnikov, A., Wang, I.-M., & Wong, P. H. (2013). Strategic Applications of Gene Expression: From Drug Discovery/Development to Bedside. *The AAPS Journal*, *15*(2), 427–437. https://doi.org/10.1208/s12248-012-9447-1

Bannoura, S. F., Uddin, Md. H., Nagasaka, M., Fazili, F., Al-Hallak, M. N., Philip, P. A., El-Rayes, B., & Azmi, A. S. (2021). Targeting KRAS in pancreatic cancer: New drugs on the horizon. *Cancer and Metastasis Reviews*, *40*(3), 819–835. https://doi.org/10.1007/s10555-021-09990-2

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets--update. *Nucleic Acids Research*, *41*(Database issue), D991-995. https://doi.org/10.1093/nar/gks1193

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y.-J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., … Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, *463*(7283), 899–905. https://doi.org/10.1038/nature08822

Bhamidipati, D., Yedururi, S., Huse, J., Chinapuvvula, S. V., Wu, J., & Subbiah, V. (2023). Exceptional Responses to Selpercatinib in RET Fusion-Driven Metastatic Pancreatic Cancer. *JCO Precision Oncology*, *7*, e2300252. https://doi.org/10.1200/PO.23.00252

Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., & Garcia-Fandino, R. (2023). The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals*, *16*(6), 891. https://doi.org/10.3390/ph16060891

Bommhardt, U., Schraven, B., & Simeoni, L. (2019). Beyond TCR Signaling: Emerging Functions of Lck in Cancer and Immunotherapy. *International Journal of Molecular Sciences*, *20*(14), 3500. https://doi.org/10.3390/ijms20143500

Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., & Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinformatics*, *20*(1), 446. https://doi.org/10.1186/s12859-019-3036-6

Campa, D., Rizzato, C., Stolzenberg-Solomon, R., Pacetti, P., Vodicka, P., Cleary, S. P., Capurso, G., Bueno-de-Mesquita, H. B. A., Werner, J., Gazouli, M., Butterbach, K., Ivanauskas, A., Giese, N., Petersen, G. M., Fogar, P., Wang, Z., Bassi, C., Ryska, M., Theodoropoulos, G. E., … Canzian, F. (2015). TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *International Journal of Cancer*, *137*(9), 2175–2183. https://doi.org/10.1002/ijc.29590

Cancer Genome Atlas Network. (2015). Genomic Classification of Cutaneous Melanoma. *Cell*, *161*(7), 1681–1696. https://doi.org/10.1016/j.cell.2015.05.044

Chang, L., Ruiz, P., Ito, T., & Sellers, W. R. (2021). Targeting pan-essential genes in cancer: Challenges and opportunities. *Cancer Cell*, *39*(4), 466–479. https://doi.org/10.1016/j.ccell.2020.12.008

Chen, J. Y., Mamidipalli, S., & Huan, T. (2009). HAPPI: An online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics*, *10*(1), S16. https://doi.org/10.1186/1471-2164-10-S1-S16

Chen, J. Y., Pandey, R., & Nguyen, T. M. (2017). HAPPI-2: A Comprehensive and High-quality Map of Human Annotated and Predicted Protein Interactions. *BMC Genomics*, *18*(1), 182. https://doi.org/10.1186/s12864-017-3512-1

Chen, J. Y., Piquette-Miller, M., & Smith, B. P. (2013). Network medicine: Finding the links to personalized therapy. *Clinical Pharmacology and Therapeutics*, *94*(6), 613–616. https://doi.org/10.1038/clpt.2013.195

Chen, J. Y., Shen, C., & Sivachenko, A. Y. (2006). Mining Alzheimer disease relevant proteins from integrated protein interactome data. *Pacific Symposium on Biocomputing*. *Pacific Symposium on Biocomputing*, 367–378.

Cheng, Y., Diao, D., Zhang, H., Song, Y.-C., & Dang, C.-X. (2013). Proliferation enhanced by NGF-NTRK1 signaling makes pancreatic cancer cells more sensitive to 2DG-induced apoptosis. *International Journal of Medical Sciences*, *10*(5), 634–640. https://doi.org/10.7150/ijms.5547

Chowbina, S. R., Wu, X., Zhang, F., Li, P. M., Pandey, R., Kasamsetty, H. N., & Chen, J. Y. (2009). HPD: An online integrated human pathway database enabling systems biology studies. *BMC Bioinformatics*, *10*(11), S5. https://doi.org/10.1186/1471-2105-10-S11-S5

Conway, J. R., Herrmann, D., Evans, T. J., Morton, J. P., & Timpson, P. (2019). Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine. *Gut*, *68*(4), 742–758. https://doi.org/10.1136/gutjnl-2018-316822

de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T. N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., Abeshouse, A., Baiceanu, D., Ciftci, E., Dogrusoz, U., Dufilie, A., Erkoc, Z., Garcia Lara, E., Fu, Z., Gross, B., … Schultz, N. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Research*, *83*(23), 3861–3867. https://doi.org/10.1158/0008-5472.CAN-23-0816

Dinstag, G., & Shamir, R. (2020). PRODIGY: Personalized prioritization of driver genes. *Bioinformatics*, *36*(6), 1831–1839. https://doi.org/10.1093/bioinformatics/btz815

Faleiro, I., Roberto, V. P., Demirkol Canli, S., Fraunhoffer, N. A., Iovanna, J., Gure, A. O., Link, W., & Castelo-Branco, P. (2021). DNA Methylation of PI3K/AKT Pathway-Related Genes Predicts Outcome in Patients with Pancreatic Cancer: A Comprehensive Bioinformatics-Based Study. *Cancers*, *13*(24), Article 24. https://doi.org/10.3390/cancers13246354

Gong, E., & Chen, J. Y. (2023). *Prioritizing Complex Disease Genes from Heterogeneous Public Databases*. Bioinformatics. https://doi.org/10.1101/2023.02.09.527562

Grassilli, S., Brugnoli, F., Lattanzio, R., Buglioni, S., & 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*(10), 379. https://doi.org/10.3390/biomedicines8100379

Gu, L., Hickey, R. J., & Malkas, L. H. (2023). Therapeutic Targeting of DNA Replication Stress in Cancer. *Genes*, *14*(7), 1346. https://doi.org/10.3390/genes14071346

Hayashi, A., Hong, J., & Iacobuzio-Donahue, C. A. (2021). The pancreatic cancer genome revisited. *Nature Reviews Gastroenterology & Hepatology*, *18*(7), 469–481. https://doi.org/10.1038/s41575-021-00463-z

Hilbig, A. (2008). Src Kinase and Pancreatic Cancer. In *Pancreatic Cancer* (Vol. 177, pp. 179–185). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71279-4_19

Hingorani, S. R., Petricoin, E. F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M. A., Ross, S., Conrads, T. P., Veenstra, T. D., Hitt, B. A., Kawaguchi, Y., Johann, D., Liotta, L. A., Crawford, H. C., Putt, M. E., Jacks, T., Wright, C. V. E., Hruban, R. H., Lowy, A. M., & Tuveson, D. A. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*, *4*(6), 437–450. https://doi.org/10.1016/s1535-6108(03)00309-x

Hu, J., Wang, J., Guo, X., Fan, Q., Li, X., Li, K., Wang, Z., Liang, S., Amin, B., Zhang, N., Chen, C., & Zhu, B. (2024). MSLN induced EMT, cancer stem cell traits and chemotherapy resistance of pancreatic cancer cells. *Heliyon*, *10*(8), e29210. https://doi.org/10.1016/j.heliyon.2024.e29210

Hu, J.-X., Zhao, C.-F., Chen, W.-B., Liu, Q.-C., Li, Q.-W., Lin, Y.-Y., & Gao, F. (2021). Pancreatic cancer: A review of epidemiology, trend, and risk factors. *World Journal of Gastroenterology*, *27*(27), 4298–4321. https://doi.org/10.3748/wjg.v27.i27.4298

Huan, T., Wu, X., & Chen, J. Y. (2010a). Systems biology visualization tools for drug target discovery. *Expert Opinion on Drug Discovery*, *5*(5), 425–439. https://doi.org/10.1517/17460441003725102

Huan, T., Wu, X., & Chen, J. Y. (2010b). Systems biology visualization tools for drug target discovery. *Expert Opinion on Drug Discovery*, *5*(5), 425–439. https://doi.org/10.1517/17460441003725102

Huang, B., Lang, X., & Li, X. (2022). The role of IL-6/JAK2/STAT3 signaling pathway in cancers. *Frontiers in Oncology*, *12*, 1023177. https://doi.org/10.3389/fonc.2022.1023177

Huang, H., Nguyen, T., Ibrahim, S., Shantharam, S., Yue, Z., & Chen, J. Y. (2015). DMAP: A connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinformatics*, *16*(S13), S4. https://doi.org/10.1186/1471-2105-16-S13-S4

Huang, H., Wu, X., Pandey, R., Li, J., Zhao, G., Ibrahim, S., & Chen, J. Y. (2012a). C2Maps: A network pharmacology database with comprehensive disease-gene-drug connectivity relationships. *BMC Genomics*, *13*(S6), S17. https://doi.org/10.1186/1471-2164-13-S6-S17

Huang, H., Wu, X., Pandey, R., Li, J., Zhao, G., Ibrahim, S., & Chen, J. Y. (2012b). C2Maps: A network pharmacology database with comprehensive disease-gene-drug connectivity relationships. *BMC Genomics*, *13*(6), S17. https://doi.org/10.1186/1471-2164-13-S6-S17

Huang, H., Wu, X., Sonachalam, M., Mandape, S. N., Pandey, R., MacDorman, K. F., Wan, P., & Chen, J. Y. (2012). PAGED: A pathway and gene-set enrichment database to enable molecular phenotype discoveries. *BMC Bioinformatics*, *13*(15), S2. https://doi.org/10.1186/1471-2105-13-S15-S2

Huang, J., Chen, P., Liu, K., Liu, J., Zhou, B., Wu, R., Peng, Q., Liu, Z. X., Li, C., Kroemer, G., Lotze, M., Zeh, H., Kang, R., & Tang, D. (2021). CDK1/2/5 inhibition overcomes IFNG-mediated adaptive immune resistance in pancreatic cancer. *Gut*, *70*(5), 890–899. https://doi.org/10.1136/gutjnl-2019-320441

Janyasupab, P., Suratanee, A., & Plaimas, K. (2021). Network diffusion with centrality measures to identify disease-related genes. *Mathematical Biosciences and Engineering*, *18*(3), 2909–2929. https://doi.org/10.3934/mbe.2021147

Jiang, H., Fu, Q., Song, X., Ge, C., Li, R., Li, Z., Zeng, B., Li, C., Wang, Y., Xue, Y., Luo, R., & Fang, W. (2019). HDGF and PRKCA upregulation is associated with a poor prognosis in patients with lung adenocarcinoma. *Oncology Letters*, *18*(5), 4936–4946. https://doi.org/10.3892/ol.2019.10812

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Science: A Publication of the Protein Society*, *28*(11), 1947–1951. https://doi.org/10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2025). KEGG: Biological systems database as a model of the real world. *Nucleic Acids Research*, *53*(D1), D672–D677. https://doi.org/10.1093/nar/gkae909

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., & 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. *Research Synthesis Methods*, *15*(4), 616–626. https://doi.org/10.1002/jrsm.1715

Lawrence, M. S., Stojanov, P., Mermel, C. H., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumor types. *Nature*, *505*(7484), 495–501. https://doi.org/10.1038/nature12912

Lazzeroni, L. C., Lu, Y., & Belitskaya-Lévy, I. (2014). P-values in genomics: Apparent precision masks high uncertainty. *Molecular Psychiatry*, *19*(12), 1336–1340. https://doi.org/10.1038/mp.2013.184

Li, J., Zhu, X., & Chen, J. Y. (2010). Discovering breast cancer drug candidates from biomedical literature. *International Journal of Data Mining and Bioinformatics*, *4*(3), 241–255. https://doi.org/10.1504/IJDMB.2010.033519

Li, W., Chen, Q., Gao, W., & Zeng, H. (2022). ARID1A promotes chemosensitivity to gemcitabine in pancreatic cancer through epigenetic silencing of RRM2. *Die Pharmazie*, *77*(7), 224–229. https://doi.org/10.1691/ph.2022.1881

Lim, B., Greer, Y., Lipkowitz, S., & Takebe, N. (2019). Novel Apoptosis-Inducing Agents for the Treatment of Cancer, a New Arsenal in the Toolbox. *Cancers*, *11*(8), 1087. https://doi.org/10.3390/cancers11081087

Lindsay, M. A. (2003). Target discovery. *Nature Reviews Drug Discovery*, *2*(10), 831–838. https://doi.org/10.1038/nrd1202

Liu, X., Qian, D., Liu, H., Abbruzzese, J. L., Luo, S., Walsh, K. M., & Wei, Q. (2020). Genetic variants of the peroxisome proliferator-activated receptor (PPAR) signaling pathway genes and risk of pancreatic cancer. *Molecular Carcinogenesis*, *59*(8), 930–939. https://doi.org/10.1002/mc.23208

Liu, Z., Roberts, R. A., Lal-Nag, M., Chen, X., Huang, R., & Tong, W. (2021). AI-based language models powering drug discovery and development. *Drug Discovery Today*, *26*(11), 2593–2607. https://doi.org/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., Pazos, A., Pérez-Castillo, Y., & Tejera, E. (2018). Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. *Scientific Reports*, *8*(1), 16679. https://doi.org/10.1038/s41598-018-35149-1

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luo, J. (2021). KRAS mutation in pancreatic cancer. *Seminars in Oncology*, *48*(1), 10–18. https://doi.org/10.1053/j.seminoncol.2021.02.003

Magger, O., Waldman, Y. Y., Ruppin, E., & Sharan, R. (2012). Enhancing the Prioritization of Disease-Causing Genes through Tissue Specific Protein Interaction Networks. *PLoS Computational Biology*, *8*(9), e1002690. https://doi.org/10.1371/journal.pcbi.1002690

Marabelle, A., Le, D. T., Ascierto, P. A., Di Giacomo, A. M., De Jesus-Acosta, A., Delord, J.-P., Geva, R., Gottfried, M., Penel, N., Hansen, A. R., Piha-Paul, S. A., Doi, T., Gao, B., Chung, H. C., Lopez-Martin, J., Bang, Y.-J., Frommer, R. S., Shah, M., Ghori, R., … Diaz, L. A. (2020). Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *38*(1), 1–10. https://doi.org/10.1200/JCO.19.02105

McCarthy, D. J., & Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics (Oxford, England)*, *25*(6), 765–771. https://doi.org/10.1093/bioinformatics/btp053

Millis, S. Z., Ikeda, S., Reddy, S., Gatalica, Z., & Kurzrock, R. (2016). Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19 784 Diverse Solid Tumors. *JAMA Oncology*, *2*(12), 1565–1573. https://doi.org/10.1001/jamaoncol.2016.0891

Mohsen, H., Gunasekharan, V., Qing, T., Seay, M., Surovtseva, Y., Negahban, S., Szallasi, Z., Pusztai, L., & Gerstein, M. B. (2021). Network propagation-based prioritization of long tail genes in 17 cancer types. *Genome Biology*, *22*(1), 287. https://doi.org/10.1186/s13059-021-02504-x

Mutch, D. M., Berger, A., Mansourian, R., Rytz, A., & Roberts, M.-A. (2002). The limit fold change model: A practical approach for selecting differentially expressed genes from microarray data. *BMC Bioinformatics*, *3*, 17. https://doi.org/10.1186/1471-2105-3-17

Oniani, D., Hilsman, J., Zang, C., Wang, J., Cai, L., Zawala, J., & Wang, Y. (2024). Emerging opportunities of using large language models for translation between drug molecules and indications. *Scientific Reports*, *14*(1), 10738. https://doi.org/10.1038/s41598-024-61124-0

Paananen, J., & Fortino, V. (2020). An omics perspective on drug target discovery platforms. *Briefings in Bioinformatics*, *21*(6), 1937–1953. https://doi.org/10.1093/bib/bbz122

Payne, S. N., Maher, M. E., Tran, N. H., Van De Hey, D. R., Foley, T. M., Yueh, A. E., Leystra, A. A., Pasch, C. A., Jeffrey, J. J., Clipson, L., Matkowskyj, K. A., & Deming, D. A. (2015). PIK3CA mutations can initiate pancreatic tumorigenesis and are targetable with PI3K inhibitors. *Oncogenesis*, *4*(10), e169–e169. https://doi.org/10.1038/oncsis.2015.28

Pei, Y.-F., Yin, X.-M., & Liu, X.-Q. (2018). TOP2A induces malignant character of pancreatic cancer through activating β-catenin signaling pathway. *Biochimica Et Biophysica Acta. Molecular Basis of Disease*, *1864*(1), 197–207. https://doi.org/10.1016/j.bbadis.2017.10.019

Penrod, N. M., & Moore, J. H. (2014). Influence networks based on coexpression improve drug target discovery for the development of novel cancer therapeutics. *BMC Systems Biology*, *8*(1), 12. https://doi.org/10.1186/1752-0509-8-12

Petti, M., Bizzarri, D., Verrienti, A., Falcone, R., & Farina, L. (2020). Connectivity Significance for Disease Gene Prioritization in an Expanding Universe. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17*(6), 2155–2161. https://doi.org/10.1109/TCBB.2019.2938512

Pluta, R., & Ułamek-Kozioł, M. (2020). Tau Protein-Targeted Therapies in Alzheimer's Disease: Current State and Future Perspectives. In X. Huang (Ed.), *Alzheimer's Disease: Drug Discovery*. Exon Publications. http://www.ncbi.nlm.nih.gov/books/NBK566118/

Poh, A. R., & Ernst, M. (2023). Functional roles of SRC signaling in pancreatic cancer: Recent insights provide novel therapeutic opportunities. *Oncogene*, *42*(22), 1786–1801. https://doi.org/10.1038/s41388-023-02701-x

Pothula, S. P., Xu, Z., Goldstein, D., Pirola, R. C., Wilson, J. S., & Apte, M. V. (2020). Targeting HGF/c-MET Axis in Pancreatic Cancer. *International Journal of Molecular Sciences*, *21*(23), 9170. https://doi.org/10.3390/ijms21239170

Rodríguez-Pérez, R., & Bajorath, J. (2021). Feature importance correlation from machine learning indicates functional relationships between proteins and similar compound binding characteristics. *Scientific Reports*, *11*(1), 14245. https://doi.org/10.1038/s41598-021-93771-y

Rosenberg, S., Simeonova, I., Bielle, F., Verreault, M., Bance, B., Le Roux, I., Daniau, M., Nadaradjane, A., Gleize, V., Paris, S., Marie, Y., Giry, M., Polivka, M., Figarella-Branger, D., Aubriot-Lorton, M.-H., Villa, C., Vasiljevic, A., Lechapt-Zalcman, E., Kalamarides, M., … Sanson, M. (2018). A recurrent point mutation in PRKCA is a hallmark of chordoid gliomas. *Nature Communications*, *9*(1), 2371. https://doi.org/10.1038/s41467-018-04622-w

Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, *616*(7958), 673–685. https://doi.org/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. *Healthcare (Basel, Switzerland)*, *11*(6), 887. https://doi.org/10.3390/healthcare11060887

Sellers, W. R., & Fisher, D. E. (1999). Apoptosis and cancer drug targeting. *Journal of Clinical Investigation*, *104*(12), 1655–1661. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC409892/

Sheng, W., Shi, X., Lin, Y., Tang, J., Jia, C., Cao, R., Sun, J., Wang, G., Zhou, L., & Dong, M. (2020). Musashi2 promotes EGF-induced EMT in pancreatic cancer via ZEB1-ERK/MAPK signaling. *Journal of Experimental & Clinical Cancer Research: CR*, *39*(1), 16. https://doi.org/10.1186/s13046-020-1521-4

Shim, J. E., Hwang, S., & Lee, I. (2015). Pathway-Dependent Effectiveness of Network Algorithms for Gene Prioritization. *PLOS ONE*, *10*(6), e0130589. https://doi.org/10.1371/journal.pone.0130589

Si, H., Zhang, N., Shi, C., Luo, Z., & Hou, S. (2023). Tumor-suppressive miR-29c binds to MAPK1 inhibiting the ERK/MAPK pathway in pancreatic cancer. *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*, *25*(3), 803–816. https://doi.org/10.1007/s12094-022-02991-9

Sigismund, S., Avanzato, D., & Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Molecular Oncology*, *12*(1), 3–20. https://doi.org/10.1002/1878-0261.12155

Singh, N., Vayer, P., Tanwar, S., Poyet, J.-L., Tsaioun, K., & Villoutreix, B. O. (2023). Drug discovery and development: Introduction to the general public and patient groups. *Frontiers in Drug Discovery*, *3*. https://doi.org/10.3389/fddsv.2023.1201419

Sinkala, M. (2023). Mutational landscape of cancer-driver genes across human cancers. *Scientific Reports*, *13*(1), 12742. https://doi.org/10.1038/s41598-023-39608-2

Sivaram, N., McLaughlin, P. A., Han, H. V., Petrenko, O., Jiang, Y.-P., Ballou, L. M., Pham, K., Liu, C., Van Der Velden, A. W. M., & Lin, R. Z. (2019). Tumor-intrinsic PIK3CA represses tumor immunogenicity in a model of pancreatic cancer. *Journal of Clinical Investigation*, *129*(8), 3264–3276. https://doi.org/10.1172/JCI123540

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational Methods in Drug Discovery. *Pharmacological Reviews*, *66*(1), 334–395. https://doi.org/10.1124/pr.112.007336

Somarelli, J. A., Boddy, A. M., Gardner, H. L., DeWitt, S. B., Tuohy, J., Megquier, K., Sheth, M. U., Hsu, S. D., Thorne, J. L., London, C. A., & Eward, W. C. (2019). Improving Cancer Drug Discovery by Studying Cancer across the Tree of Life. *Molecular Biology and Evolution*, *37*(1), 11–17. https://doi.org/10.1093/molbev/msz254

Sonehara, K., & Okada, Y. (2021). Genomics-driven drug discovery based on disease-susceptibility genes. *Inflammation and Regeneration*, *41*(1), 8. https://doi.org/10.1186/s41232-021-00158-7

Stanciu, S., Ionita-Radu, F., Stefani, C., Miricescu, D., Stanescu-Spinu, I.-I., Greabu, M., Ripszky Totan, A., & Jinga, M. (2022). Targeting PI3K/AKT/mTOR Signaling Pathway in Pancreatic Cancer: From Molecular to Clinical Aspects. *International Journal of Molecular Sciences*, *23*(17), 10132. https://doi.org/10.3390/ijms231710132

Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., Haase, K., Hames-Fathi, S., Ajayi, D., Verfaillie, A., Dhami, P., McLatchie, A., Lechner, M., Light, N., Shlien, A., Malkin, D., Feber, A., Proszek, P., Lesluyes, T., Mertens, F., … Pillay, N. (2022). Signatures of copy number alterations in human cancer. *Nature*, *606*(7916), 984–991. https://doi.org/10.1038/s41586-022-04738-6

Su, L., Chen, Y., Huang, C., Wu, S., Wang, X., Zhao, X., Xu, Q., Sun, R., Kong, X., Jiang, X., Qiu, X., Huang, X., Wang, M., & Wong, P.-P. (2023). Targeting Src reactivates pyroptosis to reverse chemoresistance in lung and pancreatic cancer models. *Science Translational Medicine*, *15*(678), eabl7895. https://doi.org/10.1126/scitranslmed.abl7895

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica. B*, *12*(7), 3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Sun, Y., Liu, Y., Ma, X., & Hu, H. (2021). The Influence of Cell Cycle Regulation on Chemotherapy. *International Journal of Molecular Sciences*, *22*(13), 6923. https://doi.org/10.3390/ijms22136923

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Tanoli, Z., Schulman ,Aron, & and Aittokallio, T. (2025). Validation guidelines for drug-target prediction methods. *Expert Opinion on Drug Discovery*, *20*(1), 31–45. https://doi.org/10.1080/17460441.2024.2430955

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., … Forbes, S. A. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

Taylor, A. J., Panzhinskiy, E., Orban, P. C., Lynn, F. C., Schaeffer, D. F., Johnson, J. D., Kopp, J. L., & Verchere, C. B. (2023). Islet amyloid polypeptide does not suppress pancreatic cancer. *Molecular Metabolism*, *68*, 101667. https://doi.org/10.1016/j.molmet.2023.101667

Timar, J., & Kashofer, K. (2020). Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Reviews*, *39*(4), 1029–1038. https://doi.org/10.1007/s10555-020-09915-5

Trajanoska, K., Bhérer, C., Taliun, D., Zhou, S., Richards, J. B., & Mooser, V. (2023). From target discovery to clinical drug development with human genetics. *Nature*, *620*(7975), 737–745. https://doi.org/10.1038/s41586-023-06388-8

Tripathi, S., Gabriel, K., Tripathi, P. K., & Kim, E. (2024). Large language models reshaping molecular biology and drug development. *Chemical Biology & Drug Design*, *103*(6), e14568. https://doi.org/10.1111/cbdd.14568

Tu, M., Klein, L., Espinet, E., Georgomanolis, T., Wegwitz, F., Li, X., Urbach, L., Danieli-Mackay, A., Küffer, S., Bojarczuk, K., Mizi, A., Günesdogan, U., Chapuy, B., Gu, Z., Neesse, A., Kishore, U., Ströbel, P., Hessmann, E., Hahn, S. A., … Singh, S. K. (2021). TNF-α-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. *Nature Cancer*, *2*(11), 1185–1203. https://doi.org/10.1038/s43018-021-00258-w

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)*, *347*(6220), 1260419. https://doi.org/10.1126/science.1260419

Van de Sande, B., Lee, J. S., Mutasa-Gottgens, E., Naughton, B., Bacon, W., Manning, J., Wang, Y., Pollard, J., Mendez, M., Hill, J., Kumar, N., Cao, X., Chen, X., Khaladkar, M., Wen, J., Leach, A., & Ferran, E. (2023). Applications of single-cell RNA sequencing in drug discovery and development. *Nature Reviews Drug Discovery*, *22*(6), 496–520. https://doi.org/10.1038/s41573-023-00688-4

Verma, H. K., Kampalli, P. K., Lakkakula, S., Chalikonda, G., Bhaskar, L. V. K. S., & Pattnaik, S. (2020). A Retrospective Look at Anti-EGFR Agents in Pancreatic Cancer Therapy. *Current Drug Metabolism*, *20*(12), 958–966. https://doi.org/10.2174/1389200220666191122104955

Villalona-Calero, M. A., Ritch, P., Figueroa, J. A., Otterson, G. A., Belt, R., Dow, E., George, S., Leonardo, J., McCachren, S., Miller, G. L., Modiano, M., Valdivieso, M., Geary, R., Oliver, J. W., & Holmlund, J. (2004). A phase I/II study of LY900003, an antisense inhibitor of protein kinase C-alpha, in combination with cisplatin and gemcitabine in patients with advanced non-small cell lung cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *10*(18 Pt 1), 6086–6093. https://doi.org/10.1158/1078-0432.CCR-04-0779

Wang, W., Li, C., Dai, Y., Wu, Q., & Yu, W. (2024). Unraveling metabolic characteristics and clinical implications in gastric cancer through single-cell resolution analysis. *Frontiers in Molecular Biosciences*, *11*. https://doi.org/10.3389/fmolb.2024.1399679

Wiedmann, L., De Angelis Rigotti, F., Vaquero-Siguero, N., Donato, E., Espinet, E., Moll, I., Alsina-Sanchis, E., Bohnenberger, H., Fernandez-Florido, E., Mülfarth, R., Vacca, M., Gerwing, J., Conradi, L.-C., Ströbel, P., Trumpp, A., Mogler, C., Fischer, A., & Rodriguez-Vita, J. (2023). HAPLN1 potentiates peritoneal metastasis in pancreatic cancer. *Nature Communications*, *14*(1), 2353. https://doi.org/10.1038/s41467-023-38064-w

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wu, F., He, J., Deng, Q., Chen, J., Peng, M., Xiao, J., Zeng, Y., Yi, L., Li, Z., Tian, R., & Jiang, Z. (2023). Neuroglobin inhibits pancreatic cancer proliferation and metastasis by targeting the GNAI1/EGFR/AKT/ERK signaling axis. *Biochemical and Biophysical Research Communications*, *664*, 108–116. https://doi.org/10.1016/j.bbrc.2023.04.080

Wu, X., Huang, H., Wei, T., Pandey, R., Reinhard, C., Li, S. D., & Chen, J. Y. (2012). Network expansion and pathway enrichment analysis towards biologically significant findings from microarrays. *Journal of Integrative Bioinformatics*, *9*(2), 213. https://doi.org/10.2390/biecoll-jib-2012-213

Xu, X., Ding, Y., Jin, J., Xu, C., Hu, W., Wu, S., Ding, G., Cheng, R., Cao, L., & Jia, S. (2023). Post-translational modification of CDK1–STAT3 signaling by fisetin suppresses pancreatic cancer stem cell properties. *Cell & Bioscience*, *13*(1), 176. https://doi.org/10.1186/s13578-023-01118-z

Yue, Z., Slominski, R., Bharti, S., & Chen, J. Y. (2022). PAGER Web APP: An Interactive, Online Gene Set and Network Interpretation Tool for Functional Genomics. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.820361

Yue, Z., Willey, C. D., Hjelmeland, A. B., & Chen, J. Y. (2019). BEERE: A web server for biomedical entity expansion, ranking and explorations. *Nucleic Acids Research*, *47*(W1), W578–W586. https://doi.org/10.1093/nar/gkz428

Yue, Z., Zheng, Q., Neylon, M. T., Yoo, M., Shin, J., Zhao, Z., Tan, A. C., & Chen, J. Y. (2018). PAGER 2.0: An update to the pathway, annotated-list and gene-signature electronic repository for Human Network Biology. *Nucleic Acids Research*, *46*(D1), D668–D676. https://doi.org/10.1093/nar/gkx1040

Zhang, G., He, P., Tan, H., Budhu, A., Gaedcke, J., Ghadimi, B. M., Ried, T., Yfantis, H. G., Lee, D. H., Maitra, A., Hanna, N., Alexander, H. R., & Hussain, S. P. (2013). Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *19*(18), 4983–4993. https://doi.org/10.1158/1078-0432.CCR-13-0209

Zhang, G., Schetter, A., He, P., Funamizu, N., Gaedcke, J., Ghadimi, B. M., Ried, T., Hassan, R., Yfantis, H. G., Lee, D. H., Lacy, C., Maitra, A., Hanna, N., Alexander, H. R., & Hussain, S. P. (2012). DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PloS One*, *7*(2), e31507. https://doi.org/10.1371/journal.pone.0031507

Zhang, H., Ferguson, A., Robertson, G., Jiang, M., Zhang, T., Sudlow, C., Smith, K., Rannikmae, K., & Wu, H. (2021). Benchmarking network-based gene prioritization methods for cerebral small vessel disease. *Briefings in Bioinformatics*, *22*(5), bbab006. https://doi.org/10.1093/bib/bbab006

Zhang, H., Sun, Y., Wang, Z., Huang, X., Tang, L., Jiang, K., & Jin, X. (2024). ZDHHC20-mediated S-palmitoylation of YTHDF3 stabilizes MYC mRNA to promote pancreatic cancer progression. *Nature Communications*, *15*(1), 4642. https://doi.org/10.1038/s41467-024-49105-3

Zhou, S.-F., & Zhong, W.-Z. (2017). Drug Design and Discovery: Principles and Applications. *Molecules*, *22*(2), Article 2. https://doi.org/10.3390/molecules22020279

Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., Qiu, Y., & Chen, Y. (2022). Therapeutic target database update 2022: Facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research*, *50*(D1), D1398–D1407. https://doi.org/10.1093/nar/gkab953

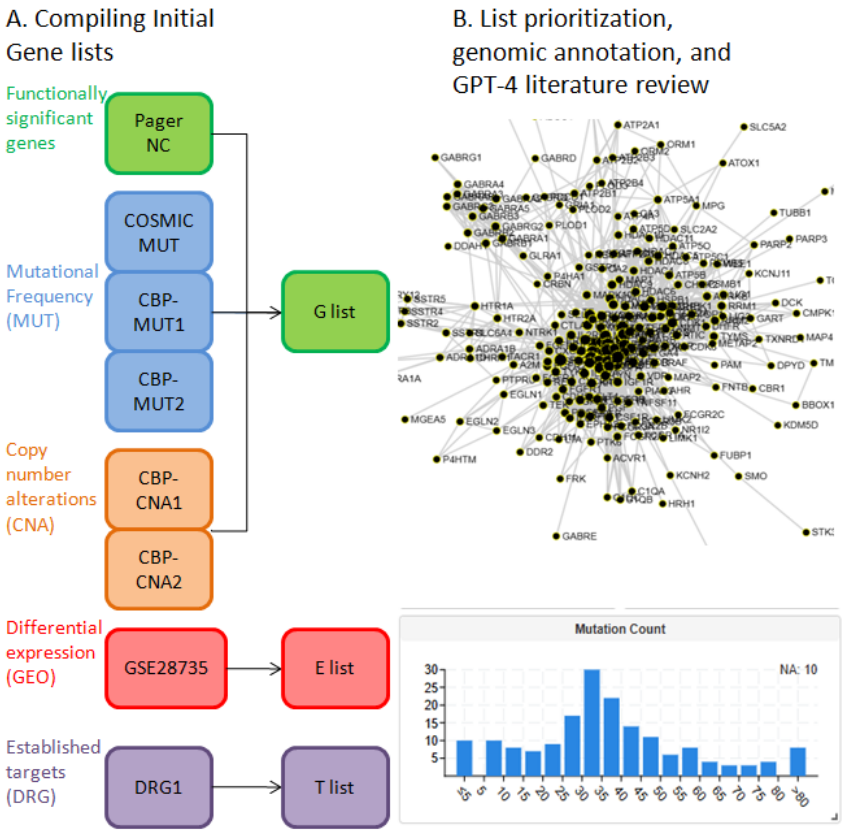
Zhu, C., Guan, X., Zhang, X., Luan, X., Song, Z., Cheng, X., Zhang, W., & Qin, J.-J. (2022). Targeting KRAS mutant cancers: From druggable therapy to drug resistance. *Molecular Cancer*, *21*(1), 159. https://doi.org/10.1186/s12943-022-01629-2

Zhu, G., Fang, Q., Zhu, F., Huang, D., & Yang, C. (2021). Structure and Function of Pancreatic Lipase-Related Protein 2 and Its Relationship With Pathological States. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.693538

Zhu, L., Jiang, M., Wang, H., Sun, H., Zhu, J., Zhao, W., Fang, Q., Yu, J., Chen, P., Wu, S., Zheng, Z., & He, Y. (2021). A narrative review of tumor heterogeneity and challenges to tumor drug therapy. *Annals of Translational Medicine*, *9*(16), 1351. https://doi.org/10.21037/atm-21-1948

Figures

Figure 1: General overview of the GET list compilation and ranking process. A. Initial gene lists from each of the three subsets are compiled. 2493 genes are compiled in the initial G list, 2000 genes are compiled in the initial E list, and 131 genes are compiled in the initial T list. B. Each list is prioritized using the BEERE network ranking and expansion tool, before being annotated with genomic information and GPT4 literature review. C. A weighted score is calculated to rank the list and genes are manually validated through literature review.



C. Ranking and validation of genes

Gene	Weighted Score	Experimentally Validated?
PIK3CA	34.2	Yes
MYC	30.1	Yes
SRC	20	Yes
EGFR	18.1	Yes
CDK1	15.9	Yes
PRKCA	15.3	Yes
TNF	12.1	Yes
LCK	11.5	Yes
JAK2	10.6	Yes
MAPK1	10.3	Yes
MTOR	9.1	Yes
AURKB	8.7	Yes
KRAS	7.7	Yes
MAPK8	7	Yes
TOP2A	6.9	Yes

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 lTCGA PanCancerAtlas)	0.079
CNA(CBIOPORTAL TCGA PANCANCERATLAS)	0.048
Mutation frequency(Cbioportal 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

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

Gene	Weighted score	CHATGPT score	GT list score	Mutation Frequency (cBioportal TCGA PanCancer Atlas)	RP-LIT score	GPT-LIT score	GET list score	Expressio 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

Figure 2: Volcano plot GSE28735: Microarray gene-expression profiles of 45 matching pairs of tumor vs. nontumor, Padj<0.05. Blue indicates down regulated while red indicates upregulated.

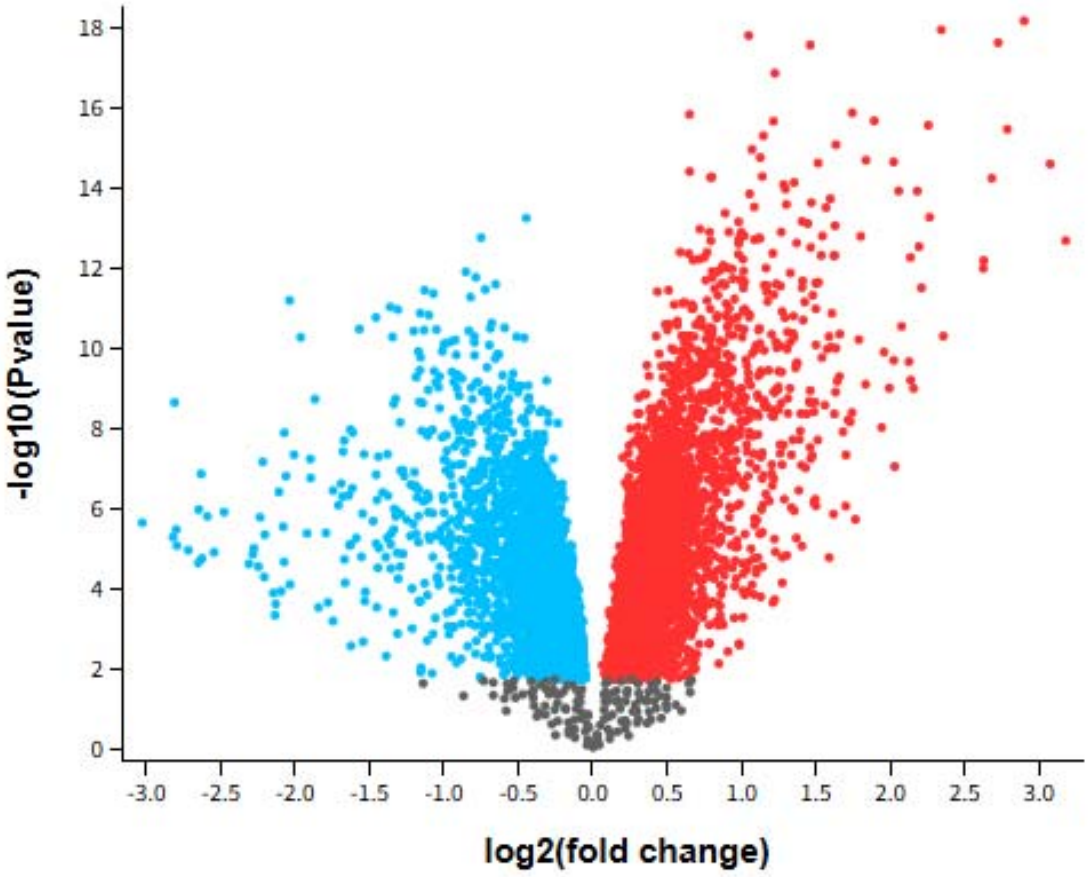


Figure 3: Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3d structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate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

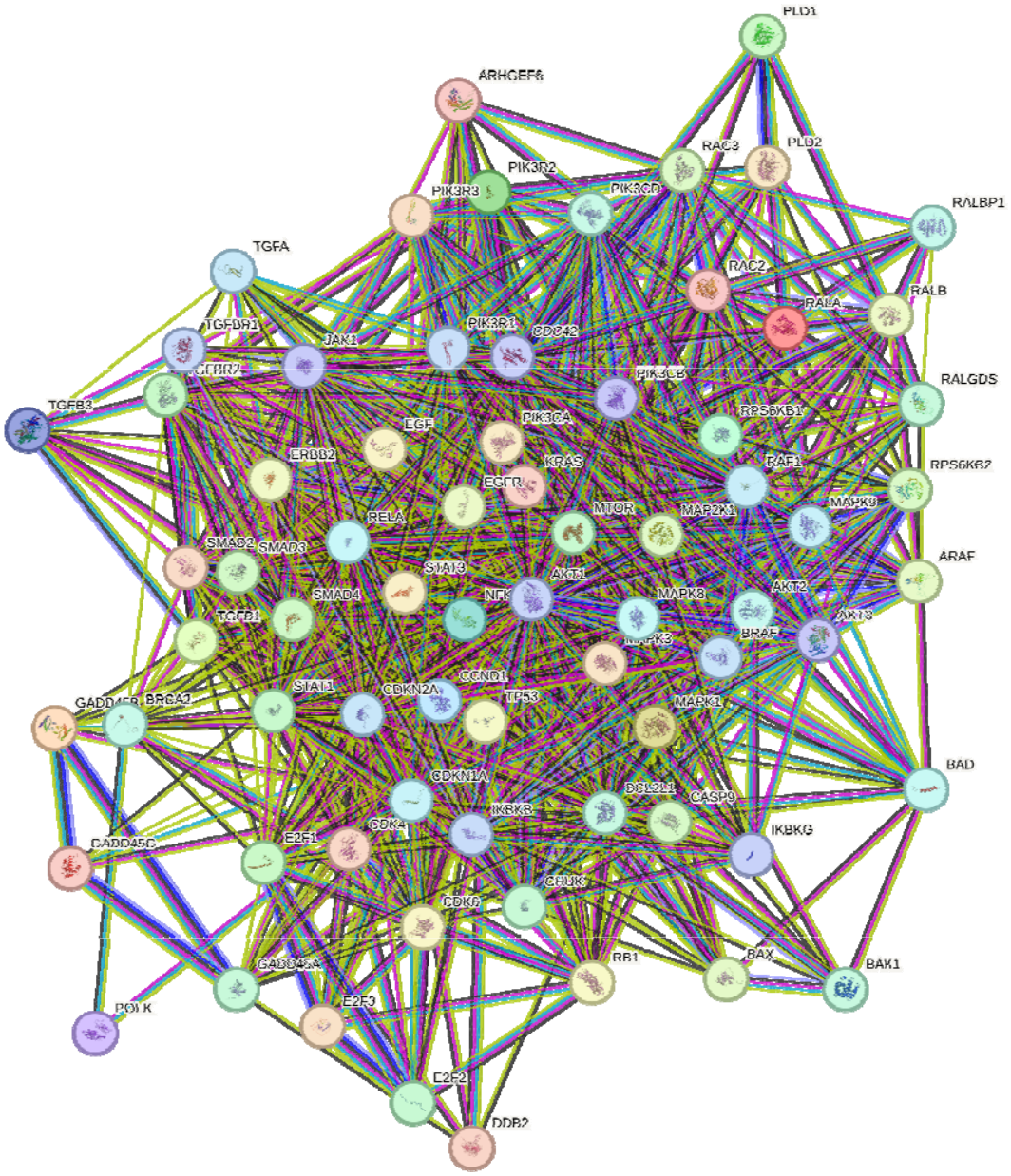


Table 3: Top 50 genes from GETGENE-AI, STRING, and GEO2R and their status as experimentally validated drug targets

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	FAM24B-CUZD1///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
ITGA4	Yes	EGF	Yes	CELA3A	No
CHEK1	Yes	MAPK3	Yes	CTRL	No
BCL2	Yes	CDC42	Yes	CTRB2///CTRB1	Yes
PRKCB	Yes	MAPK8	Yes	PNLIP///COL1A2	Yes
ERBB4	Yes	ERBB2	Yes	REG1B	Yes
RRM2	Yes	BCL2L1	Yes	REG1CP	Yes
NTRK1	Yes	MAP2K1	Yes	PDIA2	Yes
ABL1	Yes	NRAS	Yes	CPA1	Yes
KDR	Yes	PIK3R1	Yes	EGF	Yes
PTGS2	Yes	MAPK1	Yes	CELA3B	Yes
FCGR2C	No	CASP9	Yes	KIAA1324	Yes
TERT	Yes	FOXO3	Yes	NR5A2	Yes
RET	Yes	BRAF	Yes	AOX1	No
TGFB2	Yes	SMAD4	Yes	CTRB2///CTRB1	No
EGF	Yes	CDKN1B	Yes	F11	Yes
AKT1	Yes	TGFB1	Yes	ERO1B	Yes
FLT4	Yes	AKT2	Yes	KLK1	Yes
FCGR3B	Yes	BCL2L11	Yes	TMED6	No
MET	Yes	SMAD3	Yes	ANPEP	Yes
KCNA1	Yes	PIK3CB	Yes	SLC16A12	No
SSTR5	Yes	STAT1	Yes	CPB1	Yes
MSLN	Yes	CDK4	Yes	PRSS1	No

PDCD1	Yes	RAF1	No	REG1A	Yes
PTK6	Yes	SMAD2	Yes	SCGN	No
KIT	Yes	IRS1	Yes	RBPJL	Yes
ADRA1D	Yes	RELA	Yes	TRHDE	No
JAK3	Yes	RPS6KB1	No	ADH1B	No
TUBB1	Yes	AKT3	Yes	ACADL	No
CDK6	Yes	JAK1	Yes	FGL1	Yes
LIG1	Yes	E2F1	Yes	GATM	Yes
P3H3	Yes	PRKCD	Yes	PAIP2B///VPS36	Yes
LTA	Yes	CDK6	Yes	NRG4	No
KCNH2	Yes	MAPK9	Yes	NRCAM	Yes
FCGR1A	Yes	RAC2	No	FABP4	Yes
DNMT1	Yes	TGFBR2	Yes	REG3A	Yes

Figure 4: Bar graph displaying the percent of experimentally validated targets out of the top 50 genes with each framework.

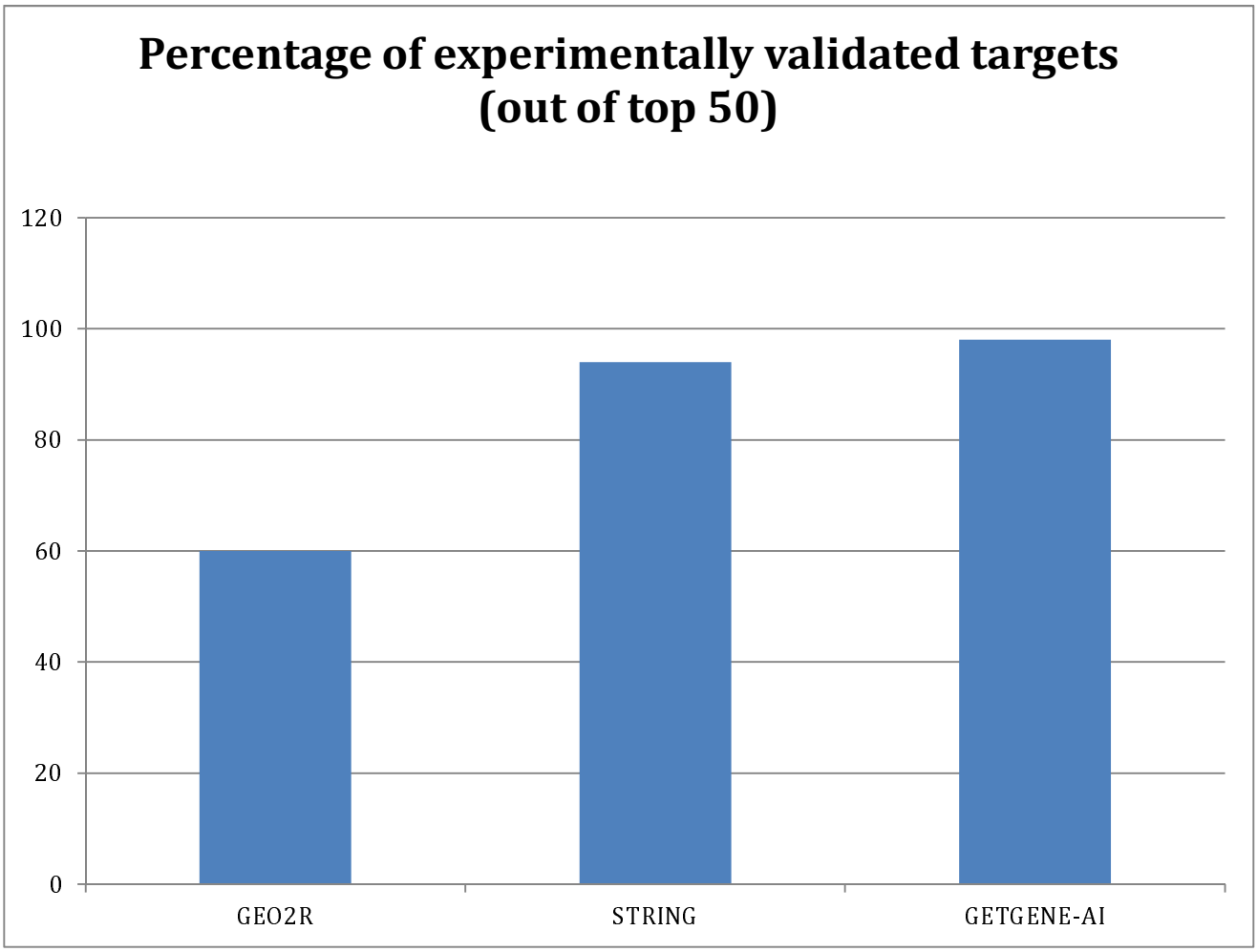
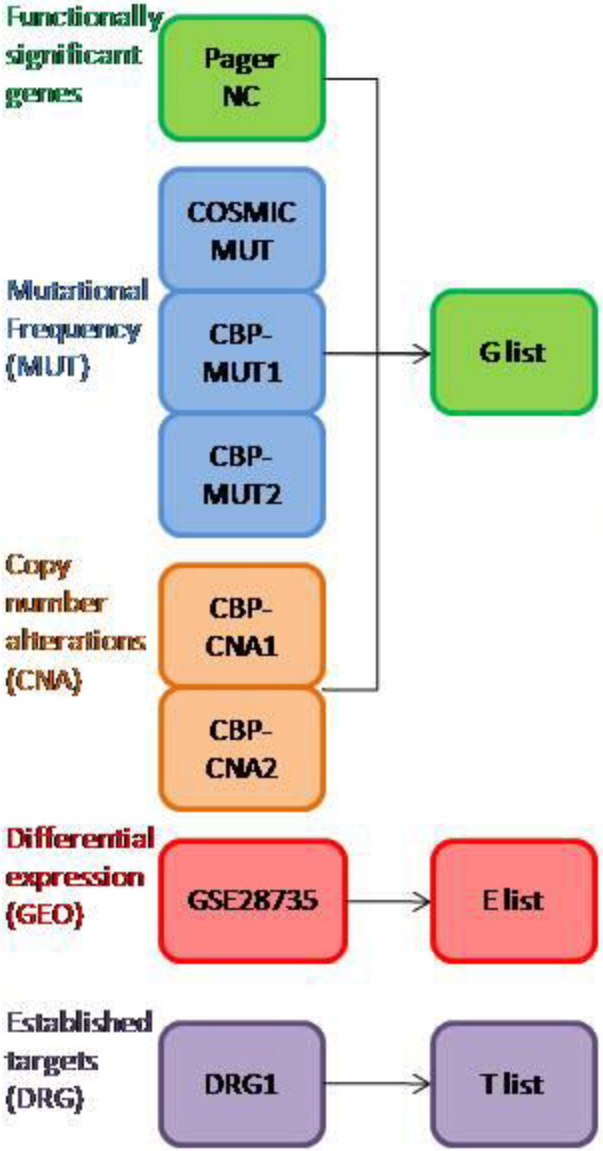


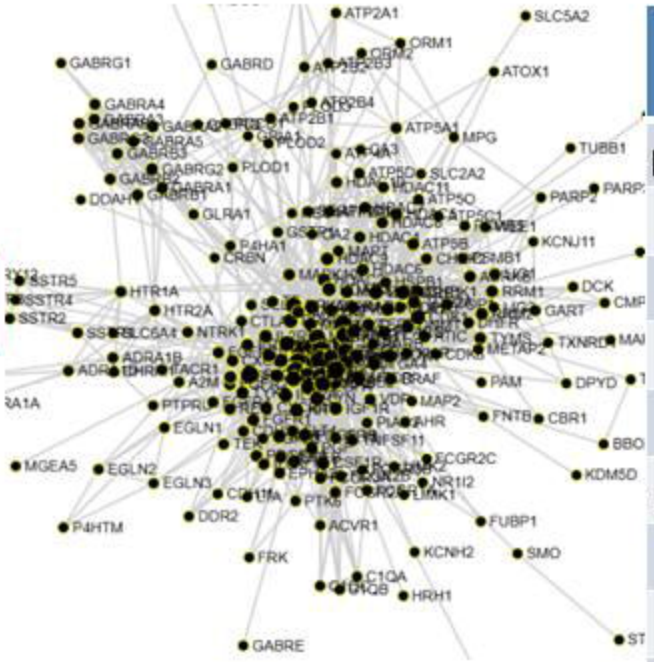
Table 4: Top 20 highest ranked genes based off of GPT4 score compared to their ranks in GET and their status as experimentally validated drug targets.

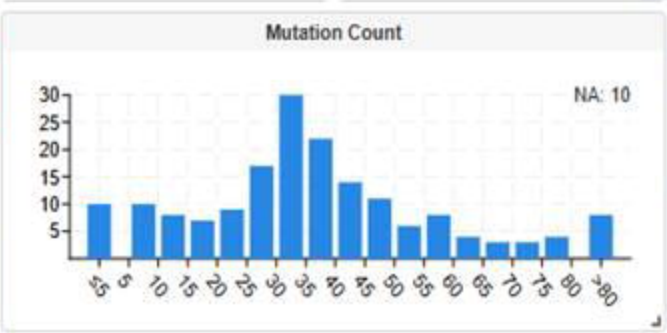
Gene	GPT4-Score ranking	GET ranking	Experimental Validation?	Citation
MYC	1	2	Yes	(Zhang et al., 2024)
SRC	2	3	Yes	(Su et al., 2023)
EGFR	3	4	Yes	(Wu et al., 2023)
TERT	4	27	Yes	(Campa et al., 2015)
RRM2	5	21	Yes	(Li et al., 2022)
PIK3CA	6	1	Yes	(Payne et al., 2015)
TOP2A	7	16	Yes	(Pei et al., 2018)
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 & 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)

A. Compiling Initial Gene lists



B. List prioritization, genomic annotation, and GPT-4 literature review





C. Ranking and validation of genes

Gene	Weighted Score	Experimentally Validated?
PIK3CA	34.2	Yes
MYC	30.1	Yes
SRC	20	Yes
EGFR	18.1	Yes
CDK1	15.9	Yes
PRKCA	15.3	Yes
TNF	12.1	Yes
LCK	11.5	Yes
JAK2	10.6	Yes
IAPK1	10.3	Yes
MTOR	9.1	Yes
URKB	8.7	Yes
KRAS	7.7	Yes
IAPK8	7	Yes
TOP2A	6.9	Yes

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 lTCGA PanCancerAtlas)	0.079
CNA(CBIOPORTAL TCGA PANCANCERATLAS)	0.048
Mutation frequency(Cbioportal 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

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

Gene	Weighted 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

Figure 2: Volcano plot GSE28735: Microarray gene-expression profiles of 45 matching pairs of tumor vs. nontumor, Padj<0.05. Blue indicates down regulated while red indicates upregulated.

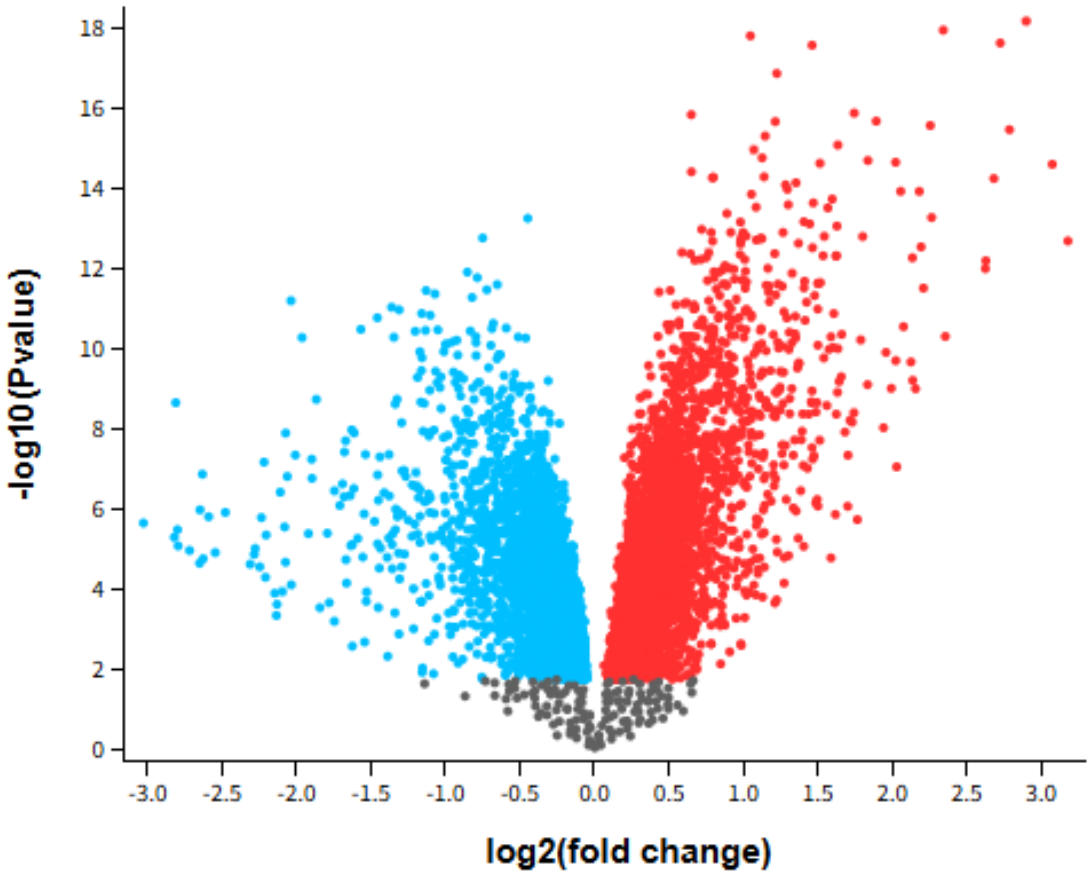


Figure 3: Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3d structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate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

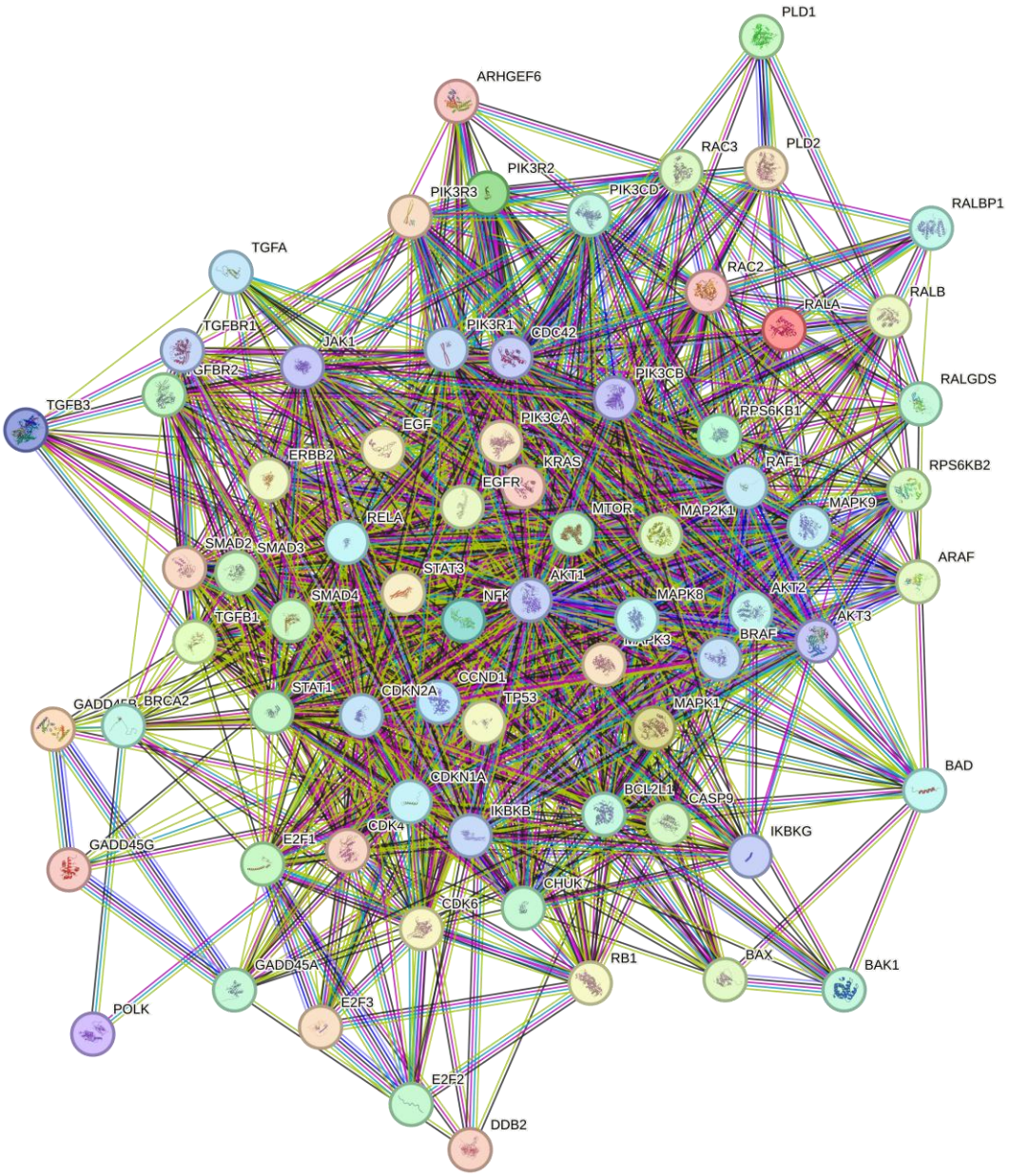


Table 3: Top 50 genes from GETGENE-AI, STRING, and GEO2R and their status as experimentally validated drug targets

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	FAM24B-CUZD1///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
ITGA4	Yes	EGF	Yes	CELA3A	No
CHEK1	Yes	MAPK3	Yes	CTRL	No
BCL2	Yes	CDC42	Yes	CTRB2///CTRB1	Yes
PRKCB	Yes	MAPK8	Yes	PNLIP///COL1A2	Yes
ERBB4	Yes	ERBB2	Yes	REG1B	Yes
RRM2	Yes	BCL2L1	Yes	REG1CP	Yes
NTRK1	Yes	MAP2K1	Yes	PDIA2	Yes
ABL1	Yes	NRAS	Yes	CPA1	Yes
KDR	Yes	PIK3R1	Yes	EGF	Yes
PTGS2	Yes	MAPK1	Yes	CELA3B	Yes
FCGR2C	No	CASP9	Yes	KIAA1324	Yes
TERT	Yes	FOXO3	Yes	NR5A2	Yes
RET	Yes	BRAF	Yes	AOX1	No
TGFB2	Yes	SMAD4	Yes	CTRB2///CTRB1	No
EGF	Yes	CDKN1B	Yes	F11	Yes
AKT1	Yes	TGFB1	Yes	ERO1B	Yes
FLT4	Yes	AKT2	Yes	KLK1	Yes
FCGR3B	Yes	BCL2L11	Yes	TMED6	No
MET	Yes	SMAD3	Yes	ANPEP	Yes
KCNA1	Yes	PIK3CB	Yes	SLC16A12	No
SSTR5	Yes	STAT1	Yes	CPB1	Yes
MSLN	Yes	CDK4	Yes	PRSS1	No

PDCD1	Yes	RAF1	No	REG1A	Yes
PTK6	Yes	SMAD2	Yes	SCGN	No
KIT	Yes	IRS1	Yes	RBPJL	Yes
ADRA1D	Yes	RELA	Yes	TRHDE	No
JAK3	Yes	RPS6KB1	No	ADH1B	No
TUBB1	Yes	AKT3	Yes	ACADL	No
CDK6	Yes	JAK1	Yes	FGL1	Yes
LIG1	Yes	E2F1	Yes	GATM	Yes
P3H3	Yes	PRKCD	Yes	PAIP2B///VPS36	Yes
LTA	Yes	CDK6	Yes	NRG4	No
KCNH2	Yes	MAPK9	Yes	NRCAM	Yes
FCGR1A	Yes	RAC2	No	FABP4	Yes
DNMT1	Yes	TGFBR2	Yes	REG3A	Yes