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

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

[1] AI.MED Lab at the Department of Biomedical Informatics and Data Science, [2] Systems Pharmacology AI Research Center, School of Medicine, the University of Alabama at Birmingham, Birmingham, AL

ABSTRACT

Identifying 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 identify 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 identified 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

Drug target discovery is a multifaceted process central to developing effective therapies for a range of diseases (S.-F. Zhou and Zhong 2017; Lindsay 2003). Traditionally, it involves the manual investigation of scientific literature and biomedical databases to identify biological targets linked to disease mechanisms, followed by evaluations based on relevance, preclinical and clinical success rates, and research popularity (Paananen and Fortino 2020; Y. Zhou et al. 2022; Trajanoska et al. 2023; Santos et al. 2017). While foundational, this approach often suffers from inefficiencies, high failure rates, and uncertainties in translating early discoveries into viable treatments (D. Sun et al. 2022; Singh et al. 2023). These challenges are particularly pronounced in cancer, where the heterogeneity of the disease demands more precise and individualized strategies (L. Zhu et al. 2021; Somarelli et al. 2019). Unlike other conditions, cancers arise from diverse genetic and molecular abnormalities, requiring advanced techniques to identify and prioritize therapeutic targets (Mroz and Rocco 2017). Cancer medications often

target rapidly dividing cells by disrupting DNA replication or cell division, as seen with chemotherapeutic agents (Gu, Hickey, and Malkas 2023; Y. Sun et al. 2021). Additionally, some therapies induce apoptosis in cancer cells by targeting pathways that regulate cell survival (Sellers and Fisher 1999; Lim et al. 2019).

Significant innovations have been developed to streamline, reform, and reduce the cost of drug target discovery, primarily through *in silico* assessments (Sadybekov and Katritch 2023). These assessments utilize neural networks, genomic datasets, and machine learning algorithms to predict key genes of interest (Serrano Nájera, Narganes Carlón, and Crowther 2021; Yue et al. 2017; Muhammad et al. 2014). These *in silico* tools offer new methods of analysis, uncovering novel insights and advancing the identification of potential therapeutic targets (Sliwoski et al. 2014; Huan, Wu, and Chen 2010). The standard *in silico* framework for drug target discovery typically begins with the construction of protein-protein interaction (PPI) networks, where potential targets are ranked based on their network connectivity and centrality (Odongo, Demiroglu-Zergeroglu, and Çakır 2024; Y. Chen and Xu 2016). Network centrality approaches to drug target discovery are commonly referred to as network-based prioritization. By leveraging network-based prioritization, researchers can efficiently analyze and interpret vast genomic datasets to identify critical regulatory genes implicated in cancer development (Chang et al. 2021; Sonehara and Okada 2021). The ability to efficiently process genomic information and derive meaningful insights is pivotal for identifying relevant drug targets, underscoring the importance of network-based prioritization (J. Y. Chen, Piquette-Miller, and Smith 2013; Huang, Li, and Chen 2009). Network-based approaches are particularly favored for this purpose, as they utilize established biological networks to identify genes associated with specific diseases (Shim, Hwang, and Lee 2015; Huang et al. 2012). These approaches prioritize disease-related genes by integrating data from PPI networks and known gene-drug associations (Mohsen et al. 2021; Zhang et al. 2021). These approaches are also able to be easily visualized (Huan, Wu, and Chen 2010). However, traditional network-based approaches often fail to incorporate crucial genomic information, such as protein expression across tissues, gene mutation frequencies, and differential gene expression profiles, which limits their utility (Petti et al. 2020). To address these limitations, supplemental *in silico* methods have been developed to achieve a more comprehensive and nuanced analysis of drug targets (Nitsch et al. 2010).

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 to calculate differential expression is by measuring "fold change," which represents the ratio of gene expression levels between these conditions (Love, Huber, and Anders 2014; Mutch et al. 2002). However, the threshold value used to define fold change can introduce arbitrariness into the prioritization process, potentially affecting the accuracy of target identification (McCarthy and Smyth 2009). GEO2R, a widely used tool, employs fold change to prioritize genes under experimental conditions, specifically comparing gene expression in cancerous versus normal tissues (Barrett et al. 2013). In addition to fold change, frequency-based prioritization methods focus on genes with higher mutational rates in disease contexts, as these genes may serve as common therapeutic targets (Dinstag and Shamir 2020; López-Cortés et al. 2018). While valuable, statistical significance methods for gene prioritization can be prone to bias, especially due to sample selection, which can skew results (Lazzeroni, Lu, and Belitskaya-Lévy 2014). To mitigate these biases, network centrality-based prioritization has been developed, leveraging gene connectivity within biological networks. This method offers a more balanced approach to gene selection, expands gene lists, and enhances disease association measures,

improving the identification of relevant therapeutic targets (Janyasupab, Suratanee, and Plaimas 2021; Magger et al. 2012). Large Language Models (LLMs), such as GPT-4, have emerged as another transformative *in silico* approach to drug discovery (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 discovery process (Khraisha et al. 2024; Li, Zhu, and Chen 2010).

To address challenges in cancer drug target discovery, we developed GETgene-AI, a framework that combines network-based analysis with artificial intelligence to prioritize actionable targets. 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 lists are derived from multiple sources, including fold change, copy number alterations, and mutational rates across various biological databases. To mitigate biases from incomplete data, GETgene-AI incorporates diverse datasets, ensuring robust prioritization. The framework iteratively refines these lists through the network-based tool BEERE, annotating genes with biological context to create a high-quality, prioritized gene list. This iterative process expands and ranks gene candidates based on biological annotations, enhancing the accuracy of target identification. By combining traditional and *in silico* methods, GETgene-AI bridges gaps in drug discovery and facilitates the development of personalized cancer therapies. Additionally, GPT-4o is integrated into the process to improve literature review efficiency and further annotate the target list, enhancing the overall workflow.

While pancreatic cancer serves as a case study in this framework, the underlying methodology is adaptable to a wide range of cancers and diseases, provided that relevant genomic and clinical data are available. This adaptability underscores the framework's potential as a versatile and powerful tool for drug discovery across diverse disease contexts. By incorporating comprehensive *in silico* analyses and leveraging high-throughput genomic data, the framework can be applied to identify novel drug targets for various malignancies, thereby accelerating the discovery of therapeutic options. Furthermore, the novel drug targets identified 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. These targets have the potential to facilitate breakthroughs in pancreatic cancer treatment, paving the way for the development of more effective, personalized therapies. Given the significant unmet clinical need in pancreatic cancer, these findings are poised to contribute to the development of targeted therapies that could improve patient outcomes and survival rates.

Methods

Our framework, "GET" (Genes, Expression, Target), provides essential biological context for gene prioritization through a comprehensive, multifaceted approach. This involves

assembling and prioritizing three distinct lists: the Gene list (G list), Expression list (E list), and Target list (T list). The G list includes genes frequently mutated, functionally significant genes, and those associated with established pathways. The E list identifies genes with the highest levels of differential expression, while the T list encompasses genes recognized as significant drug targets in the context of pancreatic cancer (PDAC). To generate these lists, disease-specific gene data were collected from diverse genomic databases. An iterative process was employed using a network-based prioritization tool to annotate each gene within biological contexts, including differential expression, protein expression in critical organs, mutational frequency, copy number alterations, and network-based ranking scores. Following the GRIPPs method for compiling statistically relevant candidate genes and iterative network-based processing (Gong and Chen 2023), modality-specific cutoffs ensured the inclusion of high-quality data in each list. Each list was then prioritized and expanded using the BEERE network-ranking tool, selected for its efficacy in refining gene lists and filtering low-quality data (Yue et al. 2019). Further annotation categorized genes based on key factors relevant to drug target prioritization, such as differential expression, mutation frequency, and copy number alterations. Candidate genes were subsequently validated through literature review, supported by large language models (e.g., ChatGPT-4). Using PDAC as a case study due to its poor prognosis (J.-X. Hu et al. 2021), our method produced quantitative data and novel insights into potential drug targets.

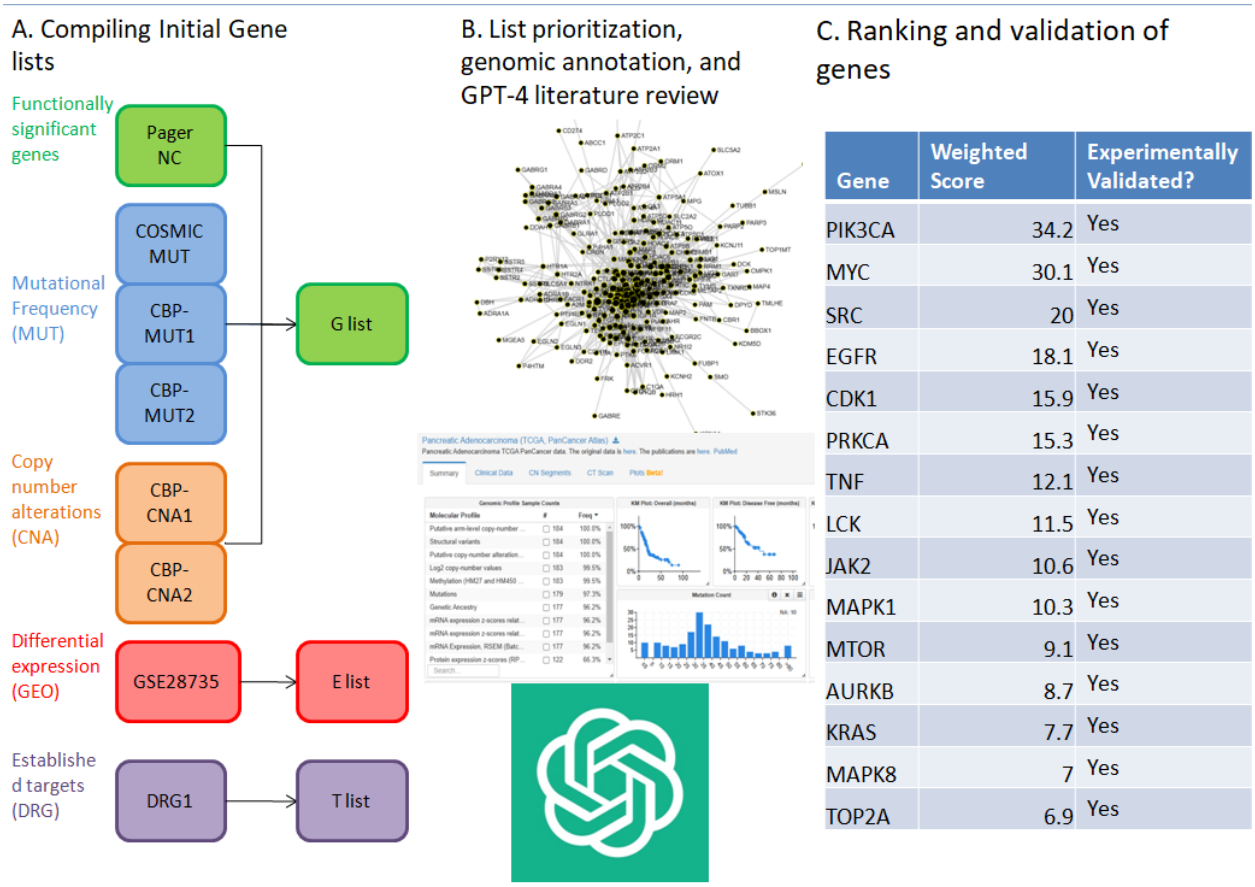
Compiling the Gene list from Genetic Mutations

For the "GENE" component of our "GET" framework, we compiled three subsets: PAGER-NC, COSMIC-MUT, and CBP-CNA-MUT. The initial "GENE" list was derived from the PAGER (Yue et al. 2018; H. Huang et al. 2012; Yue et al. 2022), cBioPortal (de Bruijn et al. 2023), and COSMIC(Tate et al. 2019) databases. To mitigate potential sample biases and data incompleteness (e.g., studies failing to detect specific genes), multiple datasets from the same databases were utilized where possible. Genes associated with the term "Pancreatic Cancer" were manually retrieved from these databases, and cutoffs were established to ensure the selection of genes most relevant to pancreatic cancer.

PAGER was employed to incorporate a biological pathway perspective into gene ranking, offering functional significance (Chowbina et al. 2009). From PAGER, 844 candidate genes were heuristically selected based on an nCoCo score ranging from 5 to 100. The nCoCo score reflects the biological relatedness of gene sets within PAGER (Yue et al. 2022; 2018; H. Huang et al. 2012). A lower cutoff of 5 was chosen because literature review indicated that functional groups below this threshold were not relevant to any disease or function, while a score above 100 was associated with major roles in other significant biological processes. For the cBioPortal and COSMIC cancer databases, cutoffs were determined based on the point at which mutational frequency no longer demonstrated significance to cancer in the literature. From cBioPortal, 1,000 genes were retrieved using a cutoff of 8.2% for copy number alterations (CNA) and 2.8% for mutational frequency. From COSMIC, 649 genes were compiled using a heuristic cutoff of 20% mutational frequency. Candidate genes from cBioPortal, PAGER, and COSMIC were then combined to form the "G list," which comprised a total of 2,493 genes.

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). BEERE expands the gene list using the

nearest-neighbor network constructed from protein-protein interactions in the HAPPI 2.0 database (X. Wu et al. 2012; Jake Yue Chen, Mamidipalli, and Huan 2009; Jake Y. Chen, Pandey, and Nguyen 2017). Following the GRIPPs method (Gong and Chen 2023) the "G list" was iteratively prioritized and expanded through BEERE, resulting in the "processed G list."



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

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 GPT4o literature review. C: A weighted score is calculated to rank the list and genes are manually validated through literature review.

Compiling candidate genes for the “Expression” subset

Candidate genes were identified through analysis performed using GEO2R on the GEO dataset GSE28735, titled "Pancreatic ductal adenocarcinoma tumor and adjacent non-tumor tissue" (G. Zhang et al. 2012; 2013). Samples were divided into tumor and non-tumor groups using the "Define groups" feature, with "human pancreatic tumor tissue patient sample" designated as the tumor group and "human pancreatic nontumor patient sample" as the non-tumor group. The dataset included a total of 90 patient samples, evenly divided into 45 tumor samples and 45 non-tumor samples. Differentially expressed genes were analyzed using the "analyze" function in GEO2R. The top 2,000 genes with the highest differential expression were compiled into the initial "E list." This list was subsequently iteratively processed using the BEERE software, following the GRIPPs method.

Compiling candidate genes for the “Target” subset

Finally, 131 genes were identified from DrugBank, a comprehensive database containing information on various drugs. Genes were extracted by searching for "Pancreatic Cancer," "Pancreatic Ductal Adenocarcinoma," and "Neuroendocrine Pancreatic Cancer" within the drugs section of DrugBank. Drugs marked as being used for the treatment of pancreatic cancer were identified by reviewing their descriptions in the Summary, Background, Indication, Associated Conditions, or Clinical Trial categories. Each drug was manually reviewed for its mechanism of action, summary, and background to determine whether it was directly used for treating pancreatic cancer or for supportive purposes, such as chemotherapy relief, pain management, or sedation. Drugs meeting these criteria had all their associated gene targets listed under the "Targets" section of DrugBank compiled. Through this process, 131 unique genes were identified.

Iterative ranking and generation of GET lists

We refined the candidate gene lists through an iterative process employing the BEERE tool for gene prioritization and expansion. A customized pipeline was developed, building upon the framework established by the GRIPPs method (Gong and Chen 2023). Using the previously compiled G, E, and T lists, we integrated the data into three distinct categories for gene prioritization. The first, the GET list, provided a comprehensive ranking by combining all three categories—mutational frequency, differential expression, and established drug targets. The second, the GT list, focused on genes that are both frequently mutated and recognized as established drug targets, offering insights into genes with high mutation frequencies and functional relevance in pancreatic cancer. Lastly, the prioritized E list was independently ranked using BEERE to specifically assess and prioritize genes based solely on their differential expression. This systematic approach ensured a robust and multifaceted analysis of candidate genes, enhancing the identification of potential therapeutic targets.

Each list underwent a standardized processing workflow. Initially, BEERE was employed to expand each list by incorporating genes from the nearest-neighbor network of protein-protein interactions derived from the HAPPI 2.0 database. This expansion enhanced the

comprehensiveness and quality of the lists. Following gene expansion, BEERE's network and significance ranking algorithms were applied to prioritize the lists, generating statistically significant rankings of candidate genes. Subsequently, each list was heuristically filtered to retain only the top 500 genes before undergoing additional rounds of refinement. This iterative process was repeated three times, as further iterations led to convergence of the lists into a single pool, diminishing their differentiation. At three iterations, the lists retained distinct gene rankings and compositions. The refined lists were then integrated to create the "Initial GET List," which underwent further prioritization and expansion via BEERE to produce the "Final GET List." Additionally, two derivative lists were constructed: the "GT List," which combined the Genes and Targets lists followed by BEERE-based expansion and prioritization, and the "Expression List," which consisted of the refined E list used for further analyses.

Annotation of GET list: Assessment of candidate genes based on clinical trials

Clinical trials play a crucial role in evaluating the efficacy of targeting specific genes. To assess the clinical relevance of each gene, clinical trial frequency was used as a metric for clinical trial popularity. Genes targeted by drugs cited in pancreatic cancer treatment were manually compiled. This process involved querying the term "Pancreatic Cancer" on ClinicalTrials.gov and collecting all drugs listed for clinical pancreatic cancer treatment. The corresponding target genes for these drugs were identified by querying the "targets" section in DrugBank. In total, 357 drugs targeting 253 genes were compiled. These genes were then annotated with BEERE scores generated through the previously described methodology. Additionally, raw, quantifiable genomic data were incorporated into the analysis. This included mutation frequency data sourced from cBioPortal (de Bruijn et al. 2023) and protein expression data from ProteinAtlas (Uhlén et al. 2015), covering key tissues such as the brain, gastrointestinal system, liver, and kidney to account for potential adverse effects of gene inhibition or activation.

GPT-4o aided literature assessment

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

The custom GPTo model was configured with specific instructions to rank genes based on a scoring system with a maximum score of 400 points, distributed across four categories: functional significance in pancreatic cancer, research popularity, treatment effectiveness when targeting or inhibiting the gene, and protein structure. Each category was allocated 100 points, and the resulting metric was termed the GPT-4 score. To mitigate GPT-4o's known issue of "hallucination" or the generation of inaccurate or nonexistent information, the model was explicitly instructed to base its rankings solely on the uploaded research database. Additionally, the model was required to cite articles referenced during the ranking process and provide

explanations for the scores assigned to each gene in every category. GPT-4 outputs were manually verified against curated datasets to ensure biological relevance and mitigate hallucinations. Citations provided by GPT-4 were cross-referenced with PubMed to confirm validity. All cited articles were manually verified, and any errors or hallucinations were addressed by instructing the model to re-search the uploaded literature database for accurate mentions of the gene.

Comprehensive gene ranking

Following the ranking of the GET list and the compilation of genes targeted in clinical trials, genes were further annotated with relevant biological information. Mutational frequency, a key category in gene ontology ranking(Timar and Kashofer 2020) ,was assessed alongside Copy Number Alterations (CNA), another critical category in ontology(Beroukhim et al. 2010) . Data for these categories were obtained from CBioPortal [30] using the “Pancreatic Cancer (UTSW, Nat Commun 2015)” and “Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)” studies, both of which utilized whole-exome sequencing for all samples.Tissue-specific expression was also considered 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.

A weighted score, termed the RP score, was developed to integrate multiple factors into the prioritization process. Spearman correlations were calculated between CNA, mutational frequencies, GET list scores, tissue expression levels, E list scores, GT list scores, and clinical trial popularity. Clinical trial popularity was defined as the number of trials testing drugs targeting specific genes for cancer treatment. This RP score provided a comprehensive and robust metric for ranking genes as potential therapeutic targets.

In Table 1, we observe the weights of each modality of ranking drug targets ordered from greatest to least.

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(cBioportal TCGA 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 1: Weights each modality was assigned for calculation of the RP score in GETgene-AI.

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.

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 sources 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

Enhancement provided by AI

GPT4o was employed to conduct a literature assessment for our gene list, though its output was not included in the final weighted score. Nonetheless, the GPT score showed strong correlations with both the weighted score and all three GET list scores. GPT-4o prioritized genes such as MYC and SRC, reflecting their prominence in the literature, which complemented GETgene-AI's reliance on network mutational analysis. To minimize false positives in the GPT-4o scoring process, we instructed GPT-4o to cite articles directly from its database. Although GPT-4o did not demonstrate a higher rate of experimental validation, it reduced the time required for the literature review by 80%. The cited articles were subsequently manually verified. RP-LIT score and GPT-score are highly correlated, with extremely similar rankings for each gene. Per Spearman correlation calculation, GPT4o's score out of 400 has a +0.457 correlation with the weighted score, indicating significant correlation.

In Table 2, we observe the ranking differences between the GPT-4 score and the GET ranking score.

Gene	GPT4-Score ranking	GET ranking	Experimental Validation?	Citation
MYC	1	2	Yes	(Huan Zhang et al. 2024)
SRC	2	3	Yes	(Su et al. 2023)
EGFR	3	4	Yes	(F. 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, Yin, and Liu 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	(J. Huang et al. 2021)
MAPK1	12	10	Yes	(Si et al. 2023)
KRAS	13	13	Yes	(Timar and Kashofer 2020)
MTOR	14	11	Yes	(Stanciu et al. 2022)
MSLN	15	37	Yes	(J. Hu et al. 2024)
RET	16	28	Yes	(Bhamidipati et al. 2023)
AKT1	17	31	Yes	(Arasanz et al. 2019)
JAK2	18	9	Yes	(B. Huang, Lang, and Li 2022)
MET	19	34	Yes	(Pothula et al. 2020)
PDCD1	20	38	Yes	(Marabelle et al. 2020)

Table 2: Top 20 highest ranked genes based off of GPT4 score compared to their ranks in GET and their status as experimentally validated drug targets.

Comparing GETgene-AI to other frameworks

We benchmarked GETgene-AI against two other frameworks: one focused on differential expression and the other on network-based prioritization. GEO2R was chosen for the differential expression comparison, using the GSE28735 dataset incorporated into the 'Expression list' aspect of our GET lists. Genes were sorted by 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 pancreatic lipase-related proteins critical for digestion and fat absorption (G. Zhu et al. 2021), but not considered viable targets for pancreatic cancer. The third-highest gene, IAPP (Islet Amyloid Polypeptide), has been shown not to function as a tumor suppressor, and loss of IAPP signaling is not linked to pancreatic cancer (Taylor et al. 2023). Of the top 50 genes in GEO2R, 30 were experimentally validated for relevance to pancreatic cancer. In comparison, GETgene-AI identified 49 experimentally validated targets in its top 50, demonstrating a 38% improvement over GEO2R. GEO2R, lacking analysis of mutational frequency, functional impact, network-based analysis, and adverse effects, falls short in drug target discovery. In contrast, GETgene-AI benefits from statistical filtering and incorporation of genomic information, improving both the efficiency and quality of genes in each list.

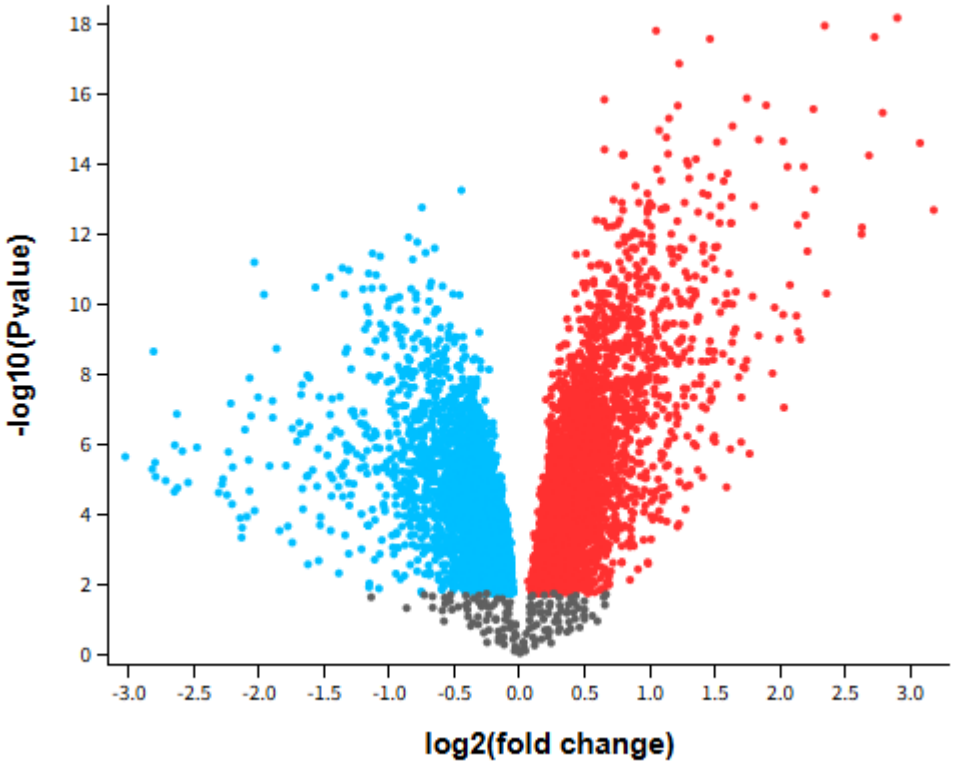


Figure 2: Volcano plot GSE28735: Microarray gene-expression profiles of 45 matching pairs of tumor vs. nontumor, Padj<0.05. Blue indicates down regulated while red indicates upregulated.

For the network-based framework, we used STRING, a database that integrates protein-protein interactions (Szklarczyk et al. 2023), and focused on the KEGG pathway hsa0512 (M. Kanehisa and Goto 2000; Minoru Kanehisa et al. 2025; Minoru Kanehisa 2019). To rank the genes, we exported the list and sorted them based on node degree, which measures 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 that stimulates cell growth and proliferation (Grassilli et al. 2020). However, AKT1 has been found to resist inhibition by switching its metabolism from glycolysis to mitochondrial cellular respiration (Arasanz et al. 2019), and it has a low mutational frequency of just 1% in a sample of 19,784 patients with various tumors (Millis et al. 2016). Due to its low mutational frequency and the challenges associated with inhibiting AKT1, it was ranked 33rd in GETgene-AI. A literature review of the top 50 genes in the STRING list revealed that 46 were experimentally validated for pancreatic cancer, whereas GETgene-AI identified 49 experimentally validated genes in the top 50, demonstrating a 6% improvement over STRING. STRING's limitations, such as its failure to account for mutational frequency and other important factors in drug target identification, result in a narrower focus, with only 81 targets identified compared to GETgene-AI's more comprehensive analysis.

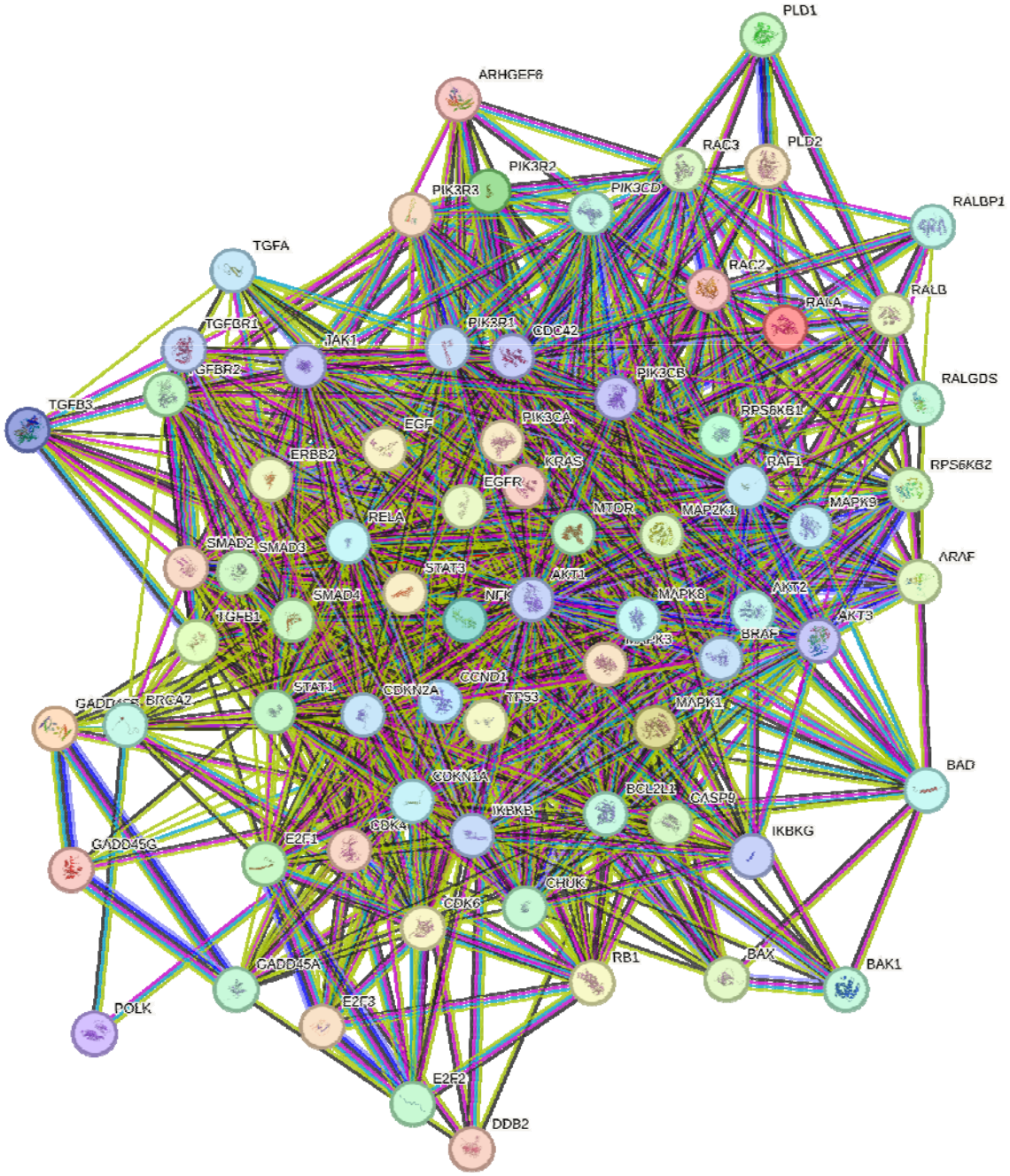


Figure 3: Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3d structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate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

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

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

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

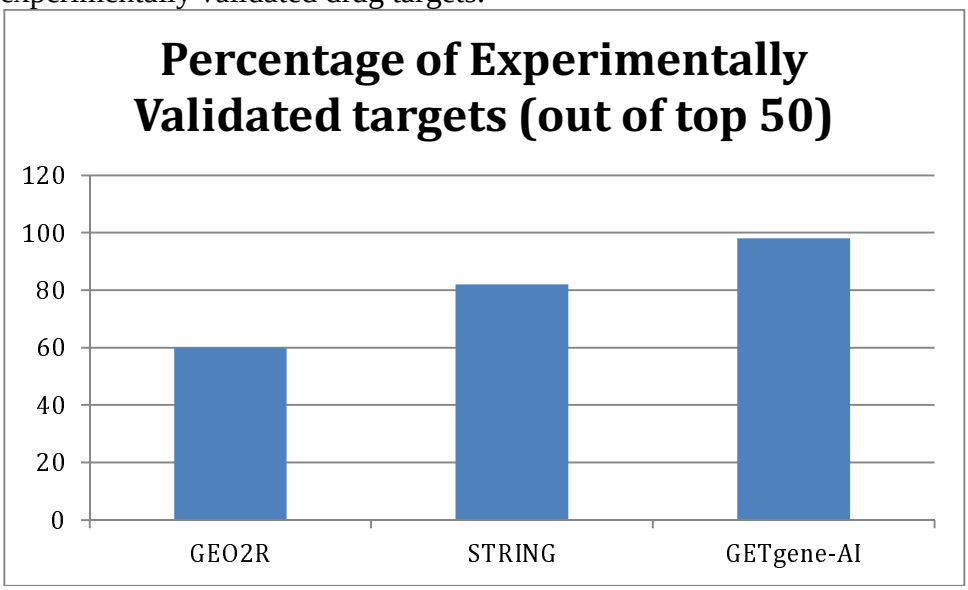


Figure 4: Bar graph displaying the percent of experimentally validated targets out of the top 50 genes with each framework.

GETgene-AI Rankings

Gene	Weighted score	CHAT GPT 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

Table 4: Highest 20 genes ranked on GETgene-AI. Weighted score is RP score, CHAT GPT score is GPT4 score.

Validation of gene rankings

During the iterative ranking process, genes with no functional relevance to cancer were systematically excluded. For example, genes ranked high due to algorithmic artifacts but without experimental validation or literature support were deprioritized.

PIK3CA ranks as the highest gene on the list. It encodes for the enzyme PI3K, which plays a crucial role in cell growth, metabolism, proliferation, and apoptosis(Conway et al. 2019). In addition, PIK3CA regulates various downstream effectors such as AKT and mTOR(Ala 2022). Mutations in PIK3CA have been shown to make cancers highly sensitive to dual therapy with PI3K/mTOR inhibitors in preclinical studies (Huayu Zhang et al. 2021), further emphasizing its potential as a key drug target. Moreover, PIK3CA-null tumors have demonstrated increased susceptibility to T-cell surveillance in vitro (Sivaram et al. 2019). In pancreatic cancer, inhibition of PIK3CA has been shown to initiate tumorigenesis in experimental studies (Payne et al. 2015), underscoring its importance in both cancer progression and therapeutic targeting.

The next highest-ranked gene on the list, MYC, is an established target in pancreatic cancer. MYC is highly ranked due to its top GET list score, which indicates its significant network centrality among the top 500 most expressed, clinically relevant, and highest mutational frequency genes. Overexpression of c-MYC is a key marker of aggressive pancreatic cancer, where it binds to the promoters of various genes (Hayashi, Hong, and Iacobuzio-Donahue 2021). Although MYC plays a crucial regulatory role in pancreatic cancer, its protein structure presents challenges for targeting. However, recent small molecule drugs have shown high efficacy in preclinical studies (Ala 2022). Despite these promising developments, MYC's inherent targeting difficulties result in a relatively low GT list score.

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

EGFR is the fourth highest ranking gene. EGFR is ranked highly due to its high GET list and Expression list scores. EGFR also plays a role in tumorigenesis, mostly in lung and breast cancer (Sigismund, Avanzato, and Lanzetti 2018). Anti-EGFR agents showed significant clinical promise, despite adverse effects (Verma et al. 2020).

KRAS ranks twelfth on our list, despite its prominence in research, with over 4,545 articles on KRAS mutations in pancreatic cancer available on PubMed. Its lower ranking is attributed to a low expression score. The KRAS oncogene plays a critical role in the initiation and maintenance of pancreatic tumors (Luo 2021). KRAS mutations are present in over 90% of pancreatic ductal adenocarcinoma (PDAC) cases, but therapeutic inhibition remains highly challenging, with 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 IFNG-triggered acquired resistance in pancreatic tumor immunity (J. 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 the development and progression of pancreatic cancer (Liu et al. 2020). Inhibition of PRKCA has shown antitumor activity in patients with advanced non-small cell lung cancer (NCSLC) (Villalona-Calero et al. 2004).

TNF was the eighth highest ranked gene on our list. TNF or Tumor Necrosis Factor upregulation is associated with invasion and immunomodulation of pancreatic cancer (Wiedmann et al. 2023). TNF mutated macrophages have also been shown to influence cancer cells into adopting more aggressive behaviors through lineage reprogramming (Tu et al. 2021).

LCK is the ninth highest ranked gene on our list. The gene LCK has been expressed in tumor cells, and is a key gene in the development of T cells (Bommhardt, Schraven, and Simeoni 2019). High LCK protein expression has been associated with improved patient survival in cancer (Cancer Genome Atlas Network 2015). The LCK gene in relation to pancreatic cancer only has four publications on PubMed as of May 2024. The identification of LCK as a high priority target demonstrates GETgene-AI's capability to identify genes with strong biological relevance but lower literature prominence.

ITGA4 is ranked 15th on our list. ITGA4 has an extremely low organ expression score. ITGA4 has only had 4 articles on PubMed discussing its role in Pancreatic Cancer. ITGA4 has potential to be an independent prognostic indicator for patient survival, and has been linked to the PI3K/AKT pathway (Faleiro et al. 2021). The identification of ITGA4 as a high priority target demonstrates GETgene-AI's capability to identify genes with strong biological relevance but lower literature prominence.

KCNA is the 34th highest ranked gene on our list. KCNA is significant, as it contains zero pubmed publications describing its relation to Pancreatic cancer. Furthermore, only three publications on PubMed have ever mentioned its relation to cancer in general. The identification of KCNA as a high priority target demonstrates GETgene-AI's capability to identify genes with strong biological relevance but lower literature prominence. Differentially high KCNA expression is observed in stomach and lung cancers, and has a positive correlation to infiltrated immune cells and survival rate (Angi et al. 2023).

False Positives and Limitations

False positives are an inherent risk in large scale computational analyses. By incorporating interative refinement and excluding genes without functional or experimental support, the GETgene-AI framework minimizes such risks. Future validation efforts will focus on further refining these rankings through experimental studies. The literature assessment provided by generative AI will also be improved as AI progresses and our model is trained on more experimental data to minimize hallucinations.

To mitigate false positives, genes with no functional relevance were systemically excluded. For example, genes ranked high due to algorithmic artifacts but without experimental validation or literature were deprioritized. Genes such as ITGA4 and PRKCB, which both contain less than 10 research articles on PubMed on its role in Pancreatic cancer were ranked lower than many popular targets due to their low scores in the GET, GT, and Expression lists as a result of deprioritizing targets without experimental validation or literature each list before RP score calculation.

This study is limited by the lack of experimental validation for the top-ranked targets. Additionally, the reliance on publicly available datasets may introduce biases due to incomplete annotations.

Broader Implications and Generalizability

The current study focuses on pancreatic cancer; however, the GETgene-AI framework can be readily adapted to other cancers or diseases with similar genomic and clinical data resources. Future studies will explore its application in breast and lung cancers by utilizing the same process described. The GETgene-AI framework's ability to implement literature review, large scale sequencing data, and network centrality scores allow for a comprehensive method of drug target prioritization. GETgene-AI is also easily scalable due to utilizing computational methods of prioritization and elimination of statistically insignificant data.

Discussion

Through our application of GETgene-AI to pancreatic cancer, we have highlighted the need to investigate potential drug targets PIK3CA, PRKCA, LCK, MAPK8, ITGA4, PRKCB and KCNA1. GETgene-AI's approach to drug target prioritization implements literature review, large scale sequencing data, network based centrality scoring, and assessment of potential adverse effects through organ expression scores. This implementation provides a scalable and comprehensive approach towards drug target prioritization which can be replicated to other cancers with similar data available. GETGENE-AI's ability to deprioritize genes with low mutational relevance demonstrates its superiority in narrowing down actionable targets efficiently.

Contributions and Limitations Provided by GPT4o

GPT4o significantly enhanced the efficiency of literature-based ranking by automating abstract reviews and prioritization. GPT4o increased the efficiency of literature review by over

80%. However, inherent challenges, such as hallucination, required manual verification to ensure accuracy. While GPT4o adds value, its integration should be approached cautiously, with safeguards to mitigate errors. Utilizing more experimental data to train GPT4o will also increase its accuracy when performing prioritization in the future.

Future Directions

The current study does not explore non-cancer applications or other disease contexts. Future studies will validate the framework in additional cancers, such as breast and lung, and assess its applicability to neurodegenerative diseases such as Alzheimer's and Parkinson's. Nonetheless, by integrating computational methods with large-scale genomic data, the GETGENE-AI framework bridges critical gaps in drug discovery, enabling faster identification of actionable targets and advancing personalized medicine.

Future work will focus on validating the top-ranked targets, such as PIK3CA and PRKCA, using CRISPR-mediated knockouts in pancreatic cancer cell lines. Drug response assays will be performed to assess the therapeutic relevance of these targets.

Conclusions

The GET framework represents a significant advancement in computational drug discovery, integrating network-based prioritization with machine learning to identify 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, Moein. 2022. "Target C-Myc to Treat Pancreatic Cancer." *Cancer Biology & Therapy* 23 (1): 34–50. https://doi.org/10.1080/15384047.2021.2017223.

Angi, Beatrice, Silvia Muccioli, Ildikò Szabò, and Luigi Leanza. 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., M. Zuazo, E. Santamaría, A. I. Bocanegra, M. Gato-Cañas, G. Fernández-Hinojal, C. Hernández-Saez, et al. 2019. "Adaption of Pancreatic Cancer Cells to AKT1 Inhibition Induces the Acquisition of Cancer Stem-Cell like Phenotype through Upregulation of Mitochondrial Functions." *Annals of Oncology* 30 (October):v11. https://doi.org/10.1093/annonc/mdz238.036.

Bannoura, Sahar F., Md. Hafiz Uddin, Misako Nagasaka, Farzeen Fazili, Mohammed Najeeb Al-Hallak, Philip A. Philip, Bassel El-Rayes, and Asfar S. Azmi. 2021. "Targeting KRAS in Pancreatic Cancer: New Drugs on the Horizon." *Cancer and Metastasis Reviews* 40 (3): 819–35. https://doi.org/10.1007/s10555-021-09990-2.

Beroukhim, Rameen, Craig H. Mermel, Dale Porter, Guo Wei, Soumya Raychaudhuri, Jerry Donovan, Jordi Barretina, et al. 2010. "The Landscape of Somatic Copy-Number Alteration across Human Cancers." *Nature* 463 (7283): 899–905. https://doi.org/10.1038/nature08822.

Bhamidipati, Deepak, Sireesha Yedururi, Jason Huse, Sri Veda Chinapuvvula, Jie Wu, and Vivek Subbiah. 2023. "Exceptional Responses to Selpercatinib in RET Fusion-Driven Metastatic Pancreatic Cancer." *JCO Precision Oncology* 7 (September):e2300252. https://doi.org/10.1200/PO.23.00252.

Bozhilova, Lyuba V., Alan V. Whitmore, Jonny Wray, Gesine Reinert, and Charlotte M. Deane. 2019. "Measuring Rank Robustness in Scored Protein Interaction Networks." *BMC Bioinformatics* 20 (1): 446. https://doi.org/10.1186/s12859-019-3036-6.

Bruijn, Ino de, Ritika Kundra, Brooke Mastrogiacomo, Thinh Ngoc Tran, Luke Sikina, Tali Mazor, Xiang Li, et al. 2023. "Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal." *Cancer Research* 83 (23): 3861–67. https://doi.org/10.1158/0008-5472.CAN-23-0816.

Campa, Daniele, Cosmeri Rizzato, Rachael Stolzenberg-Solomon, Paola Pacetti, Pavel Vodicka, Sean P. Cleary, Gabriele Capurso, et al. 2015. "TERT Gene Harbors Multiple Variants Associated with Pancreatic Cancer Susceptibility." *International Journal of Cancer* 137 (9): 2175–83. https://doi.org/10.1002/ijc.29590.

Cancer Genome Atlas Network. 2015. "Genomic Classification of Cutaneous Melanoma." *Cell* 161 (7): 1681–96. https://doi.org/10.1016/j.cell.2015.05.044.

Chen, Jake Y., Ragini Pandey, and Thanh M. Nguyen. 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, Jake Yue, SudhaRani Mamidipalli, and Tianxiao Huan. 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.

Cheng, Yao, Dong-mei Diao, Hao Zhang, Yong-Chun Song, and Cheng-Xue Dang. 2013. “Proliferation Enhanced by NGF-NTRK1 Signaling Makes Pancreatic Cancer Cells More Sensitive to 2DG-Induced Apoptosis.” *International Journal of Medical Sciences* 10 (5): 634–40. https://doi.org/10.7150/ijms.5547.

Chowbina, Sudhir R., Xiaogang Wu, Fan Zhang, Peter M. Li, Ragini Pandey, Harini N. Kasamsetty, and Jake Y. Chen. 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, James Rw, David Herrmann, Tr Jeffry Evans, Jennifer P. Morton, and Paul Timpson. 2019. “Combating Pancreatic Cancer with PI3K Pathway Inhibitors in the Era of Personalised Medicine.” *Gut* 68 (4): 742–58. https://doi.org/10.1136/gutjnl-2018-316822.

Faleiro, Inês, Vânia Palma Roberto, Secil Demirkol Canli, Nicolas A. Fraunhoffer, Juan Iovanna, Ali Osmay Gure, Wolfgang Link, and Pedro Castelo-Branco. 2021. “DNA Methylation of PI3K/AKT Pathway-Related Genes Predicts Outcome in Patients with Pancreatic Cancer: A Comprehensive Bioinformatics-Based Study.” *Cancers* 13 (24): 6354. https://doi.org/10.3390/cancers13246354.

Gong, Eric, and Jake Y. Chen. 2023. “Prioritizing Complex Disease Genes from Heterogeneous Public Databases.” Bioinformatics. https://doi.org/10.1101/2023.02.09.527562.

Grassilli, Silvia, Federica Brugnoli, Rossano Lattanzio, Simonetta Buglioni, and Valeria Bertagnolo. 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, Long, Robert J. Hickey, and Linda H. Malkas. 2023. “Therapeutic Targeting of DNA Replication Stress in Cancer.” *Genes* 14 (7): 1346. https://doi.org/10.3390/genes14071346.

Hayashi, Akimasa, Jungeui Hong, and Christine A. Iacobuzio-Donahue. 2021. “The Pancreatic Cancer Genome Revisited.” *Nature Reviews Gastroenterology & Hepatology* 18 (7): 469–81. https://doi.org/10.1038/s41575-021-00463-z.

Hilbig, Andreas. 2008. “Src Kinase and Pancreatic Cancer.” In *Pancreatic Cancer*, 177:179–85. Recent Results in Cancer Research. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71279-4_19.

Hingorani, Sunil R., Emanuel F. Petricoin, Anirban Maitra, Vinodh Rajapakse, Catrina King, Michael A. Jacobetz, Sally Ross, et al. 2003. “Preinvasive and Invasive Ductal Pancreatic Cancer and Its Early Detection in the Mouse.” *Cancer Cell* 4 (6): 437–50. https://doi.org/10.1016/s1535-6108(03)00309-x.

Hu, Jian-Xiong, Cheng-Fei Zhao, Wen-Biao Chen, Qi-Cai Liu, Qu-Wen Li, Yan-Ya Lin, and Feng Gao. 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.

Hu, Jili, Jia Wang, Xu Guo, Qing Fan, Xinming Li, Kai Li, Zhuoyin Wang, et al. 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.

Huan, Tianxiao, Xiaogang Wu, and Jake Y. Chen. 2010. "Systems Biology Visualization Tools for Drug Target Discovery." *Expert Opinion on Drug Discovery* 5 (5): 425–39. https://doi.org/10.1517/17460441003725102.

Huang, Bei, Xiaoling Lang, and Xihong Li. 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, Hui, Xiaogang Wu, Madhankumar Sonachalam, Sammed N. Mandape, Ragini Pandey, Karl F. MacDorman, Ping Wan, and Jake Y. Chen. 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, Jin, Pan Chen, Ke Liu, Jiao Liu, Borong Zhou, Runliu Wu, Qiu Peng, et al. 2021. "CDK1/2/5 Inhibition Overcomes IFNG-Mediated Adaptive Immune Resistance in Pancreatic Cancer." *Gut* 70 (5): 890–99. https://doi.org/10.1136/gutjnl-2019-320441.

Jiang, Honghong, Qiaofen Fu, Xin Song, Chunlei Ge, Ruilei Li, Zhen Li, Baozhen Zeng, et al. 2019. "HDGF and PRKCA Upregulation Is Associated with a Poor Prognosis in Patients with Lung Adenocarcinoma." *Oncology Letters* 18 (5): 4936–46. https://doi.org/10.3892/ol.2019.10812.

Kanehisa, M., and S. Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28 (1): 27–30. https://doi.org/10.1093/nar/28.1.27.

Kanehisa, Minoru. 2019. "Toward Understanding the Origin and Evolution of Cellular Organisms." *Protein Science: A Publication of the Protein Society* 28 (11): 1947–51. https://doi.org/10.1002/pro.3715.

Kanehisa, Minoru, Miho Furumichi, Yoko Sato, Yuriko Matsuura, and Mari Ishiguro-Watanabe. 2025. "KEGG: Biological Systems Database as a Model of the Real World." *Nucleic Acids Research* 53 (D1): D672–77. https://doi.org/10.1093/nar/gkae909.

Khraisha, Qusai, Sophie Put, Johanna Kappenberg, Azza Warraitch, and Kristin Hadfield. 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): 616–26. https://doi.org/10.1002/jrsm.1715.

Li, Wenyi, Qiwei Chen, Weiwei Gao, and Hui Zeng. 2022. "ARID1A Promotes Chemosensitivity to Gemcitabine in Pancreatic Cancer through Epigenetic Silencing of RRM2." *Die Pharmazie* 77 (7): 224–29. https://doi.org/10.1691/ph.2022.1881.

Lim, Bora, Yoshimi Greer, Stanley Lipkowitz, and Naoko Takebe. 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, Mark A. 2003. "Target Discovery." *Nature Reviews Drug Discovery* 2 (10): 831–38. https://doi.org/10.1038/nrd1202.

Liu, Xiaowen, Danwen Qian, Hongliang Liu, James L. Abbruzzese, Sheng Luo, Kyle M. Walsh, and Qingyi Wei. 2020. "Genetic Variants of the Peroxisome Proliferator-Activated Receptor (PPAR) Signaling Pathway Genes and Risk of Pancreatic Cancer." *Molecular Carcinogenesis* 59 (8): 930–39. https://doi.org/10.1002/mc.23208.

Luo, Ji. 2021. “KRAS Mutation in Pancreatic Cancer.” *Seminars in Oncology* 48 (1): 10–18. https://doi.org/10.1053/j.seminoncol.2021.02.003.

Marabelle, Aurelien, Dung T. Le, Paolo A. Ascierto, Anna Maria Di Giacomo, Ana De Jesus-Acosta, Jean-Pierre Delord, Ravit Geva, et al. 2020. “Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study.” *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.

Millis, Sherri Z., Sadakatsu Ikeda, Sandeep Reddy, Zoran Gatalica, and Razelle Kurzrock. 2016. “Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19□784 Diverse Solid Tumors.” *JAMA Oncology* 2 (12): 1565–73. https://doi.org/10.1001/jamaoncol.2016.0891.

Mroz, Edmund A., and James W. Rocco. 2017. “The Challenges of Tumor Genetic Diversity.” *Cancer* 123 (6): 917–27. https://doi.org/10.1002/cncr.30430.

Paananen, Jussi, and Vittorio Fortino. 2020. “An Omics Perspective on Drug Target Discovery Platforms.” *Briefings in Bioinformatics* 21 (6): 1937–53. https://doi.org/10.1093/bib/bbz122.

Payne, S. N., M. E. Maher, N. H. Tran, D. R. Van De Hey, T. M. Foley, A. E. Yueh, A. A. Leystra, et al. 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, Yao-Fei, Xi-Min Yin, and Xi-Qiang Liu. 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.

Poh, Ashleigh R., and Matthias Ernst. 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, Srinivasa P., Zhihong Xu, David Goldstein, Romano C. Pirola, Jeremy S. Wilson, and Minoti V. Apte. 2020. “Targeting HGF/c-MET Axis in Pancreatic Cancer.” *International Journal of Molecular Sciences* 21 (23): 9170. https://doi.org/10.3390/ijms21239170.

Rosenberg, Shai, Iva Simeonova, Franck Bielle, Maite Verreault, Bertille Bance, Isabelle Le Roux, Mailys Daniau, et al. 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.

Santos, Rita, Oleg Ursu, Anna Gaulton, A. Patrícia Bento, Ramesh S. Donadi, Cristian G. Bologa, Anneli Karlsson, et al. 2017. “A Comprehensive Map of Molecular Drug Targets.” *Nature Reviews. Drug Discovery* 16 (1): 19–34. https://doi.org/10.1038/nrd.2016.230.

Sellers, William R., and David E. Fisher. 1999. “Apoptosis and Cancer Drug Targeting.” *Journal of Clinical Investigation* 104 (12): 1655–61. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC409892/.

Sheng, Weiwei, Xiaoyang Shi, Yiheng Lin, Jingtong Tang, Chao Jia, Rongxian Cao, Jian Sun, Guosen Wang, Lei Zhou, and Ming Dong. 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.

Si, Hongtao, Ning Zhang, Chang Shi, Zhanjiang Luo, and Senlin Hou. 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): 803–16. https://doi.org/10.1007/s12094-022-02991-9.

Sigismund, Sara, Daniele Avanzato, and Letizia Lanzetti. 2018. "Emerging Functions of the EGFR in Cancer." *Molecular Oncology* 12 (1): 3–20. https://doi.org/10.1002/1878-0261.12155.

Singh, Natesh, Philippe Vayer, Shivalika Tanwar, Jean-Luc Poyet, Katya Tsaioun, and Bruno O. Villoutreix. 2023. "Drug Discovery and Development: Introduction to the General Public and Patient Groups." *Frontiers in Drug Discovery* 3 (May). https://doi.org/10.3389/fddsv.2023.1201419.

Sivaram, Nithya, Patrick A. McLaughlin, Han V. Han, Oleksi Petrenko, Ya-Ping Jiang, Lisa M. Ballou, Kien Pham, Chen Liu, Adrianus W.M. Van Der Velden, and Richard Z. Lin. 2019. "Tumor-Intrinsic PIK3CA Represses Tumor Immunogenicity in a Model of Pancreatic Cancer." *Journal of Clinical Investigation* 129 (8): 3264–76. https://doi.org/10.1172/JCI123540.

Somarelli, Jason A, Amy M Boddy, Heather L Gardner, Suzanne Bartholf DeWitt, Joanne Tuohy, Kate Megquier, Maya U Sheth, et al. 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.

Stanciu, Silviu, Florentina Ionita-Radu, Constantin Stefani, Daniela Miricescu, Iulia-Ioana Stanescu-Spinu, Maria Greabu, Alexandra Ripszky Totan, and Mariana Jinga. 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.

Su, Liangping, Yitian Chen, Cheng Huang, Sangqing Wu, XiaoJuan Wang, Xinbao Zhao, Qiuping Xu, et al. 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, Duxin, Wei Gao, Hongxiang Hu, and Simon Zhou. 2022. "Why 90% of Clinical Drug Development Fails and How to Improve It?" *Acta Pharmaceutica Sinica. B* 12 (7): 3049–62. https://doi.org/10.1016/j.apsb.2022.02.002.

Sun, Ying, Yang Liu, Xiaoli Ma, and Hao Hu. 2021. "The Influence of Cell Cycle Regulation on Chemotherapy." *International Journal of Molecular Sciences* 22 (13): 6923. https://doi.org/10.3390/ijms22136923.

Szklarczyk, Damian, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L. Gable, et al. 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): D638–46. https://doi.org/10.1093/nar/gkac1000.

Tate, John G., Sally Bamford, Harry C. Jubb, Zbyslaw Sondka, David M. Beare, Nidhi Bindal, Harry Boutselakis, et al. 2019. "COSMIC: The Catalogue Of Somatic Mutations In Cancer." *Nucleic Acids Research* 47 (D1): D941–47. https://doi.org/10.1093/nar/gky1015.

Taylor, Austin J., Evgeniy Panzhinskiy, Paul C. Orban, Francis C. Lynn, David F. Schaeffer, James D. Johnson, Janel L. Kopp, and C. Bruce Verchere. 2023. “Islet Amyloid Polypeptide Does Not Suppress Pancreatic Cancer.” *Molecular Metabolism* 68 (January):101667. https://doi.org/10.1016/j.molmet.2023.101667.

Timar, Jozsef, and Karl Kashofer. 2020. “Molecular Epidemiology and Diagnostics of KRAS Mutations in Human Cancer.” *Cancer Metastasis Reviews* 39 (4): 1029–38. https://doi.org/10.1007/s10555-020-09915-5.

Trajanoska, Katerina, Claude Bhérer, Daniel Taliun, Sirui Zhou, J. Brent Richards, and Vincent Mooser. 2023. “From Target Discovery to Clinical Drug Development with Human Genetics.” *Nature* 620 (7975): 737–45. https://doi.org/10.1038/s41586-023-06388-8.

Tu, Mengyu, Lukas Klein, Elisa Espinet, Theodoros Georgomanolis, Florian Wegwitz, Xiaojuan Li, Laura Urbach, et al. 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, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. “Proteomics. Tissue-Based Map of the Human Proteome.” *Science (New York, N.Y.)* 347 (6220): 1260419. https://doi.org/10.1126/science.1260419.

Verma, Henu K., Praveen K. Kampalli, Saikrishna Lakkakula, Gayathri Chalikonda, Lakkakula V.K.S. Bhaskar, and Smaranika Pattnaik. 2020. “A Retrospective Look at Anti-EGFR Agents in Pancreatic Cancer Therapy.” *Current Drug Metabolism* 20 (12): 958–66. https://doi.org/10.2174/1389200220666191122104955.

Villalona-Calero, Miguel A., Paul Ritch, Jose A. Figueroa, Gregory A. Otterson, Robert Belt, Edward Dow, Sebastian George, et al. 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): 6086–93. https://doi.org/10.1158/1078-0432.CCR-04-0779.

Wiedmann, Lena, Francesca De Angelis Rigotti, Nuria Vaquero-Siguero, Elisa Donato, Elisa Espinet, Iris Moll, Elisenda Alsina-Sanchis, et al. 2023. “HAPLN1 Potentiates Peritoneal Metastasis in Pancreatic Cancer.” *Nature Communications* 14 (1): 2353. https://doi.org/10.1038/s41467-023-38064-w.

Wu, Fan, Jin He, Qianxi Deng, Jun Chen, Mingyu Peng, Jiayi Xiao, Yiwei Zeng, et al. 2023. “Neuroglobin Inhibits Pancreatic Cancer Proliferation and Metastasis by Targeting the GNAI1/EGFR/AKT/ERK Signaling Axis.” *Biochemical and Biophysical Research Communications* 664 (July):108–16. https://doi.org/10.1016/j.bbrc.2023.04.080.

Wu, Xiaogang, Hui Huang, Tao Wei, Ragini Pandey, Christoph Reinhard, Shuyu D. Li, and Jake Y. Chen. 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, Xiaodong, Yimin Ding, Junbin Jin, Chengjie Xu, Wenyi Hu, Songtao Wu, Guoping Ding, Rui Cheng, Liping Cao, and Shengnan Jia. 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, Zongliang, Radomir Slominski, Samuel Bharti, and Jake Y. Chen. 2022. "PAGER Web APP: An Interactive, Online Gene Set and Network Interpretation Tool for Functional Genomics." *Frontiers in Genetics* 13 (April). https://doi.org/10.3389/fgene.2022.820361.

Yue, Zongliang, Christopher D Willey, Anita B Hjelmeland, and Jake Y Chen. 2019. "BEERE: A Web Server for Biomedical Entity Expansion, Ranking and Explorations." *Nucleic Acids Research* 47 (W1): W578–86. https://doi.org/10.1093/nar/gkz428.

Yue, Zongliang, Qi Zheng, Michael T Neylon, Minjae Yoo, Jimin Shin, Zhiying Zhao, Aik Choon Tan, and Jake Y Chen. 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): D668–76. https://doi.org/10.1093/nar/gkx1040.

Zhang, Geng, Peijun He, Hanson Tan, Anuradha Budhu, Jochen Gaedcke, B. Michael Ghadimi, Thomas Ried, et al. 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): 4983–93. https://doi.org/10.1158/1078-0432.CCR-13-0209.

Zhang, Geng, Aaron Schetter, Peijun He, Naotake Funamizu, Jochen Gaedcke, B. Michael Ghadimi, Thomas Ried, et al. 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, Huan, Yan Sun, Zhaokai Wang, Xiaoju Huang, Lu Tang, Ke Jiang, and Xin Jin. 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.

Zhang, Huayu, Amy Ferguson, Grant Robertson, Muchen Jiang, Teng Zhang, Cathie Sudlow, Keith Smith, Kristiina Rannikmae, and Honghan Wu. 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.

Zhou, Shu-Feng, and Wei-Zhu Zhong. 2017. "Drug Design and Discovery: Principles and Applications." *Molecules* 22 (2): 279. https://doi.org/10.3390/molecules22020279.

Zhou, Ying, Yintao Zhang, Xichen Lian, Fengcheng Li, Chaoxin Wang, Feng Zhu, Yunqing Qiu, and Yuzong Chen. 2022. "Therapeutic Target Database Update 2022: Facilitating Drug Discovery with Enriched Comparative Data of Targeted Agents." *Nucleic Acids Research* 50 (D1): D1398–1407. https://doi.org/10.1093/nar/gkab953.

Zhu, Guoying, Qing Fang, Fengshang Zhu, Dongping Huang, and Changqing Yang. 2021. "Structure and Function of Pancreatic Lipase-Related Protein 2 and Its Relationship With Pathological States." *Frontiers in Genetics* 12 (July). https://doi.org/10.3389/fgene.2021.693538.

Zhu, Liang, Minlin Jiang, Hao Wang, Hui Sun, Jun Zhu, Wencheng Zhao, Qiyu Fang, et al. 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.

Zhang, Huan et al. "ZDHHC20-mediated S-palmitoylation of YTHDF3 stabilizes MYC mRNA to promote pancreatic cancer progression." Nature communications vol. 15,1 4642. 31 May. 2024, doi:10.1038/s41467-024-49105-3

Wu, Fan et al. "Neuroglobin inhibits pancreatic cancer proliferation and metastasis by targeting

the GNAI1/EGFR/AKT/ERK signaling axis." Biochemical and biophysical research communications vol. 664 (2023): 108-116. doi:10.1016/j.bbrc.2023.04.080

Campa, Daniele et al. "TERT gene harbors multiple variants associated with pancreatic cancer susceptibility." International journal of cancer vol. 137,9 (2015): 2175-83. doi:10.1002/ijc.29590

Li, Wenyi et al. "ARID1A promotes chemosensitivity to gemcitabine in pancreatic cancer through epigenetic silencing of RRM2." Die Pharmazie vol. 77,7 (2022): 224-229. doi:10.1691/ph.2022.1881

Pei, Yao-Fei et al. "TOP2A induces malignant character of pancreatic cancer through activating β-catenin signaling pathway." Biochimica et biophysica acta. Molecular basis of disease vol. 1864,1 (2018): 197-207. doi:10.1016/j.bbadis.2017.10.019

Cheng, Yao et al. "Proliferation enhanced by NGF-NTRK1 signaling makes pancreatic cancer cells more sensitive to 2DG-induced apoptosis." International journal of medical sciences vol. 10,5 (2013): 634-40. doi:10.7150/ijms.5547

Hingorani, Sunil R et al. "Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse." Cancer cell vol. 4,6 (2003): 437-50. doi:10.1016/s1535-6108(03)00309-x

Sheng, Weiwei et al. "Musashi2 promotes EGF-induced EMT in pancreatic cancer via ZEB1-ERK/MAPK signaling." Journal of experimental & clinical cancer research : CR vol. 39,1 16. 17 Jan. 2020, doi:10.1186/s13046-020-1521-4

Si, Hongtao et al. "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 vol. 25,3 (2023): 803-816. doi:10.1007/s12094-022-02991-9

Stanciu, Silviu et al. "Targeting PI3K/AKT/mTOR Signaling Pathway in Pancreatic Cancer: From Molecular to Clinical Aspects." International journal of molecular sciences vol. 23,17 10132. 4 Sep. 2022, doi:10.3390/ijms231710132

Hu, Jili et al. "MSLN induced EMT, cancer stem cell traits and chemotherapy resistance of pancreatic cancer cells." Heliyon vol. 10,8 e29210. 7 Apr. 2024, doi:10.1016/j.heliyon.2024.e29210

Bhamidipati, Deepak et al. "Exceptional Responses to Selpercatinib in RET Fusion-Driven Metastatic Pancreatic Cancer." JCO precision oncology vol. 7 (2023): e2300252. doi:10.1200/PO.23.00252

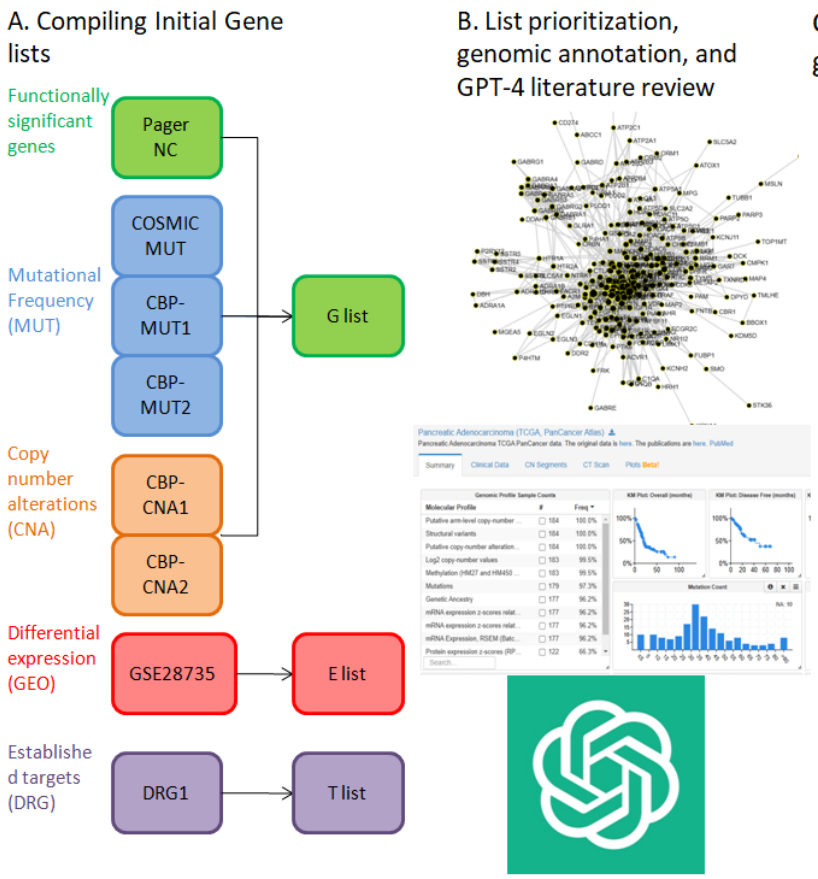
Huang, Bei et al. "The role of IL-6/JAK2/STAT3 signaling pathway in cancers." Frontiers in oncology vol. 12 1023177. 16 Dec. 2022, doi:10.3389/fonc.2022.1023177

Pothula, Srinivasa P et al. "Targeting HGF/c-MET Axis in Pancreatic Cancer." International journal of molecular sciences vol. 21,23 9170. 1 Dec. 2020, doi:10.3390/ijms21239170

Marabelle, Aurelien et al. "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 vol. 38,1 (2020): 1-10. doi:10.1200/JCO.19.02105

Figures

Figure 1



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

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.

Table 1:

In Table 1, we observe the weights of each modality of ranking drug targets ordered from greatest to least.

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(cBioportal TCGA 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 1: Weights each modality was assigned for calculation of the RP score in GETGENE-AI.

Table 2:

Gene	GPT4-Score ranking	GET ranking	Experimental Validation?	Citation
MYC	1	2	Yes	(Huan Zhang et al. 2024)
SRC	2	3	Yes	(Su et al. 2023)
EGFR	3	4	Yes	(F. 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, Yin, and Liu 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	(J. Huang et al. 2021)
MAPK1	12	10	Yes	(Si et al. 2023)
KRAS	13	13	Yes	(Timar and Kashofer 2020)
MTOR	14	11	Yes	(Stanciu et al. 2022)
MSLN	15	37	Yes	(J. Hu et al. 2024)
RET	16	28	Yes	(Bhamidipati et al. 2023)
AKT1	17	31	Yes	(Arasanz et al. 2019)
JAK2	18	9	Yes	(B. Huang, Lang, and Li 2022)
MET	19	34	Yes	(Pothula et al. 2020)
PDCD1	20	38	Yes	(Marabelle et al. 2020)

Table 2: Top 20 highest ranked genes based off of GPT4 score compared to their ranks in GET and their status as experimentally validated drug targets.

Figure 2:

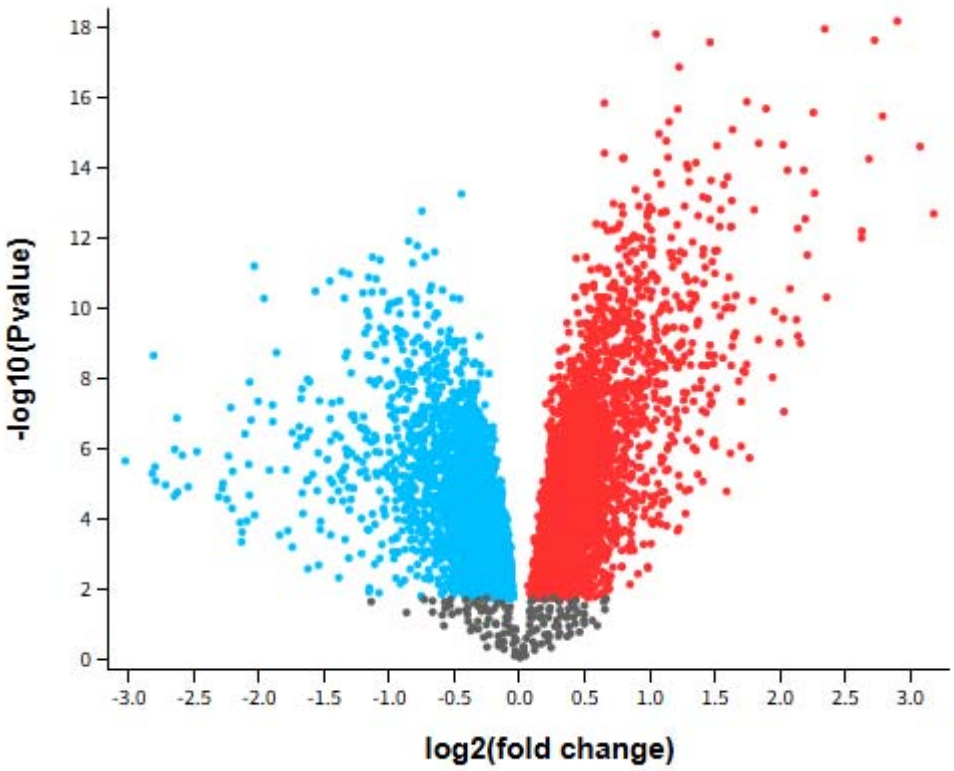


Figure 2: Figure 2: Volcano plot GSE28735: Microarray gene-expression profiles of 45 matching pairs of tumor vs. nontumor, Padj<0.05. Blue indicates down regulated while red indicates upregulated.

Figure 3:

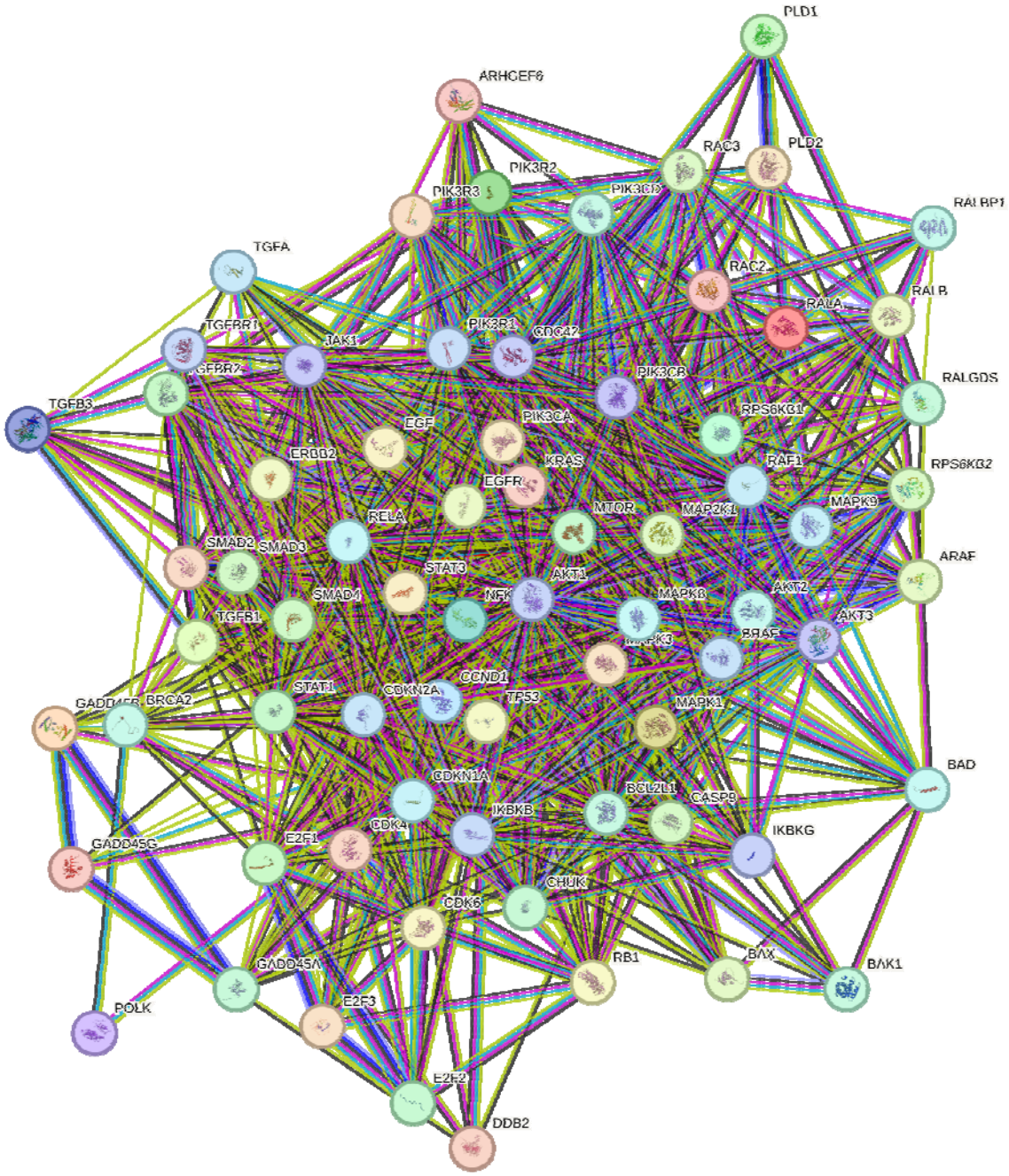


Figure 3: Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3d structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate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

represent text mining, black represents co-expression, and light purple represents protein homology.

Table 3:

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

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

Figure 4:

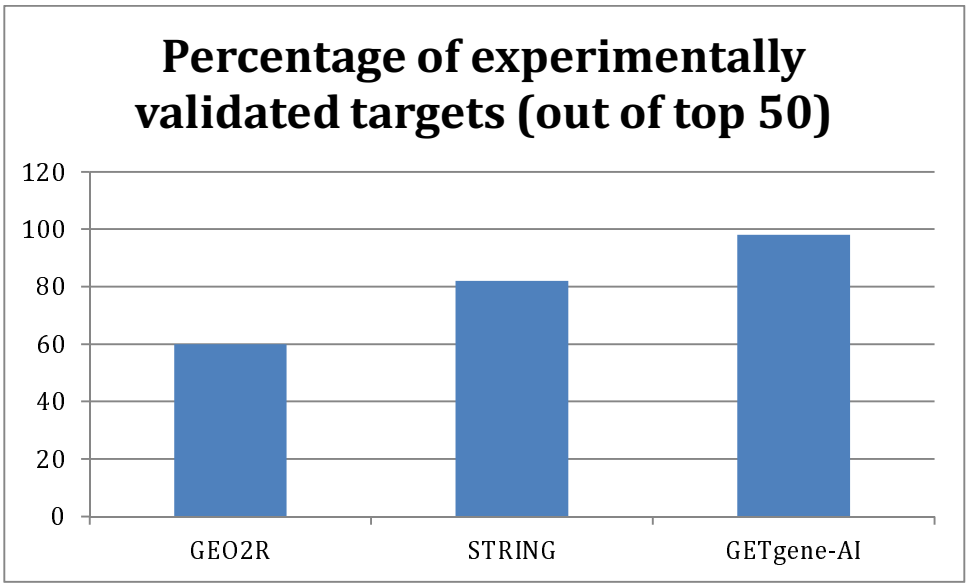


Figure 4: Bar graph displaying the percent of experimentally validated targets out of the top 50 genes with each framework.

Table 4:

Gene	Weighted score	CHAT GPT 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

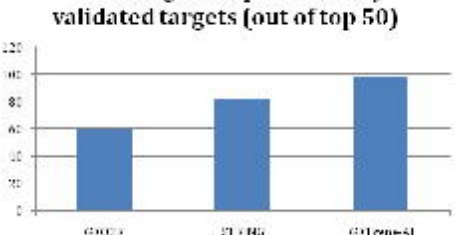
Table 4: Highest 20 genes ranked on GETGENE-AI. Weighted score is RP score, CHAT GPT score is GPT4 score.

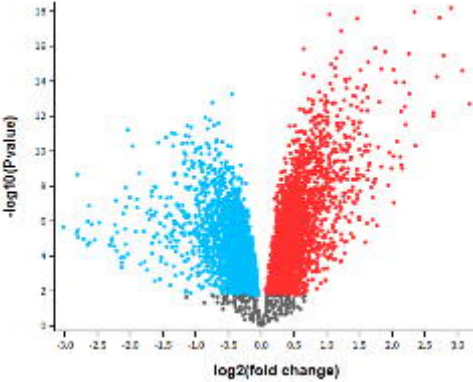
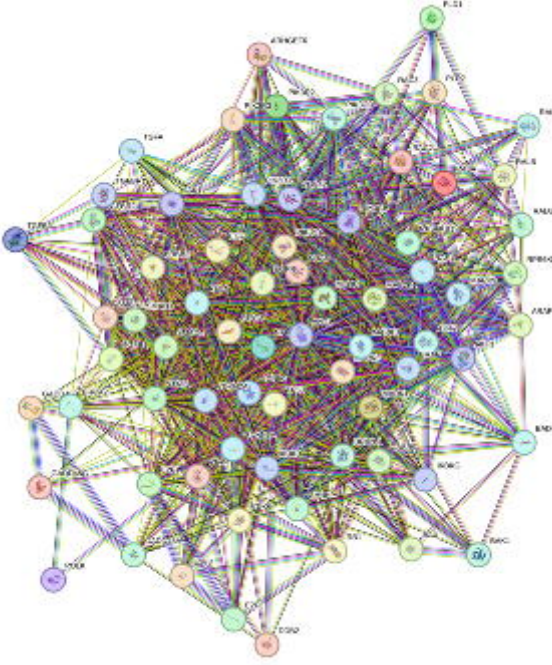
A. Compiling Initial Gene Lists

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
PIK3CA	15.3	Yes
TNF	12.1	Yes
LCK	11.5	Yes
IAK2	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


Percentage of experimentally validated targets (out of top 50)


-log10(Pvalue)
log2(fold change)