NAR Genomics and Bioinformatics, 2024, **6**, lqae157
https://doi.org/10.1093/nargab/lqae157
Advance access publication date: 21 November 2024
Standard Article

CancerPro: deciphering the pan-cancer prognostic landscape through combinatorial enrichment analysis and knowledge network insights

Zhigang Wang[1,*], Yize Yuan[1], Zhe Wang[1], Wenjia Zhang[1], Chong Chen[2], Zhaojun Duan[2], Suyuan Peng[3], Jie Zheng[4], Yongqun He[4] and Xiaolin Yang[1,*]

[1]Department of Biomedical Engineering, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China
[2]Department of Immunology, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China
[3]Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing, China
[4]Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[*]To whom correspondence should be addressed. Email: wangzg@pumc.edu.cn
Correspondence may also be addressed to Xiaolin Yang. Email: yangxl@pumc.edu.cn

Abstract

Gene expression levels serve as valuable markers for assessing prognosis in cancer patients. To understand the mechanisms underlying prognosis and explore potential therapeutics across diverse cancers, we developed CancerPro (https:/medcode.link/cancerpro). This knowledge network platform integrates comprehensive biomedical data on genes, drugs, diseases and pathways, along with their interactions. By integrating ontology and knowledge graph technologies, CancerPro offers a user-friendly interface for analyzing pan-cancer prognostic markers and exploring genes or drugs of interest. CancerPro implements three core functions: gene set enrichment analysis based on multiple annotations; in-depth drug analysis; and in-depth gene list analysis. Using CancerPro, we categorized genes and cancers into distinct groups and utilized network analysis to identify key biological pathways associated with unfavorable prognostic genes. The platform further pinpoints potential drug targets and explores potential links between prognostic markers and patient characteristics such as glutathione levels and obesity. For renal and prostate cancer, CancerPro identified risk genes linked to immune deficiency pathways and alternative splicing abnormalities. This research highlights CancerPro's potential as a valuable tool for researchers to explore pan-cancer prognostic markers and uncover novel therapeutic avenues. Its flexible tools support a wide range of biological investigations, making it a versatile asset in cancer research and beyond.

Introduction

Gene expression is a promising marker for cancer prognosis. Many studies have investigated the relationship between gene expression levels and patient survival, and these results provide guidance for personalized treatment (1). Gene expression patterns allow researchers to classify tumor samples into various molecular subtypes. These subtypes may exhibit different responses to identical treatments (2). Furthermore, alterations in gene expression can directly impact patient responses to specific treatment regimens. Therefore, researchers can analyze expression patterns to develop personalized treatment plans for individual patients. A study combined gene expression levels to predict endocrine therapy efficacy in breast cancer, identifying patients with poor treatment responses (3). Gene expression levels can be used to infer the infiltration status of immune cells in tumor microenvironments (4). The infiltration of immune cells also impacts patient prognosis. Therefore, considering the above comprehensive information, gene expression emerges as a promising marker for cancer prognosis, facilitating the study of genetic-level mechanisms influencing patient outcomes.

In cancer research, variations across different types of tumors introduce complexities and challenges to the study. A single gene can be mutated in different types of cancer, but its functions and mechanisms can differ significantly. Additionally, certain genes play dual roles, promoting some cancers while suppressing others (5). For instance, MYC, a transcription factor and an oncogene, serves as a potent driver in many human cancers, regulating multiple biological activities that promote tumorigenesis. Studies have shown that higher MYC expression is associated with metastasis and poor disease-free survival (6). Nevertheless, some studies indicate that MYC may also exert tumor-suppressive functions, and low MYC protein expression may predict poor outcomes after surgery for hepatocellular carcinoma (7). This feature of dual roles introduces additional complexity to cancer research and therapy, and it needs deeper investigation and understanding.

Therefore, conducting pan-cancer analysis to identify commonalities and unique characteristics among various tumors is

Received: March 25, 2024. **Revised:** August 26, 2024. **Editorial Decision:** October 28, 2024. **Accepted:** October 30, 2024

essential. Pan-cancer research enables the discovery of genes shared by multiple tumors as well as those specific to individual tumors through the classification of human cancer-associated genes. This approach provides valuable insights into tumor occurrence mechanisms and guides the development of effective treatment strategies.

Due to the vast number of genes associated with cancer, different researchers may focus on different genes and drugs depending on their research goals. Therefore, developing a flexible knowledge discovery platform is essential. The results of a 'census' of cancer genes by Futreal *et al.* indicate that mutations in > 1% of genes contribute to cancer in humans (8). Moreover, the molecular mechanisms and potential effects of most drugs may not be fully understood, and researchers may uncover clues indicating the therapeutic potential of certain drugs. For any given study, researchers may identify a set of genes associated with tumorigenesis or prognosis. Research on these genes or drugs can provide clues to tumor mechanisms, prognosis and potential drug discovery. Only an online knowledge insight platform can meet the research requirements for such constantly evolving conditions.

In the development process of a knowledge insight platform, both ontologies and knowledge graphs are essential. An ontology can standardize the hierarchical relationship between concepts and it serves as a formalized knowledge representation system describing relationships between concepts and their attributes. For instance, in gene function research, Gene Ontology (GO) is used to establish a standardized semantic model for genes and their functions (9). An ontology can help integrate data from multiple sources and form a comprehensive knowledge platform. Knowledge graphs are graphical representations of relationships between entities. For example, WikiPathways is a knowledge graph containing pathway information that integrates relationships among genes, proteins and metabolites in biological pathways (10). Using ontologies in knowledge graphs allows us to extract entity interaction information at different levels, and then to mine that knowledge flexibly. Researchers can utilize this flexibility to discover new associations, predict gene functions and understand biological interactions.

By combining ontologies with knowledge graphs, their respective strengths are leveraged: ontologies provide rigorous concept definitions and hierarchical structures, while knowledge graphs provide flexible representations of relationships between entities, providing a more specific understanding of biological systems.

In our study, we constructed a knowledge network database using ontologies, which we then used to develop the CancerPro Knowledge Network Insight Platform. This platform provides a comprehensive and user-friendly interface for researchers for in-depth exploration and analysis of pan-cancer prognosis genes. Our analysis grouped prognosis marker genes for 17 types of tumors. We examined gene characteristics, the involved pathways, gene expression-regulating drugs and other shared or unique information across all cancers. Through grouping tumors according to marker genes, we investigated disease characteristics and tumor-specific associated genes, and analyzed their potential mechanisms.

Materials and methods

We aimed to determine the differences and shared features of prognostic genes in multiple cancers, as well as their underlying biological mechanisms, and classify tumors based on their prognostic features to identify tumor-specific features. Therefore, data about the relationship between gene expression and prognosis were collected, and a clustering algorithm was employed to group tumors based on the expression of prognostic genes. Prognostic genes were compared with oncogenes and tumor suppressor genes to identify their differences and overlaps, and to explore their different roles in cancer development. In order to facilitate pan-cancer research, we constructed a knowledge graph by collecting a lot of information on genes, proteins, drugs, phenotypes and diseases. We developed a user-friendly interface to implement three key functions: enrichment analysis of gene sets based on multiple annotation information; in-depth analysis of drugs; and analysis of gene lists. Through these three functions, we will be able to gain a comprehensive understanding of the landscape of prognostic genes at the pan-cancer level, as well as establish a basis for future research, such as the development of new drugs. A diagram of the overall process of this study can be seen in Figure 1.

The pan-cancer gene expression and prognosis association data were obtained from The Human Protein Atlas (HPA, version 23.0) project (11). The HPA investigated mRNA expression levels of protein-coding genes across 17 major cancer types. Log-rank *P*-values were computed for Kaplan–Meier analysis to correlate mRNA expression levels with patient survival. Genes were divided into prognostic favorable, unprognostic favorable, unprognostic unfavorable and unprognostic unfavorable categories. A correlation was found between shorter patient survival and up-regulation of genes associated with cell proliferation and down-regulation of genes associated with cell differentiation.

Our primary focus was on the prognostic unfavorable and prognostic favorable gene categories. A pan-cancer prognosis gene expression matrix was constructed by assigning a value of 1 to genes classified as prognostic favorable, –1 for genes classified as prognostic unfavorable and 0 for genes without corresponding values. We selected favorable and unfavorable genes associated with at least three tumors as potential pan-cancer prognostic markers. These gene subsets were used for subsequent clustering analysis. Hierarchical clustering analysis was performed on the matrix using Euclidean distance and complete linkage. The resulting hierarchical clustering tree was grouped into a specific number of categories using the cutree function from the R language stats package. In our study, samples and genes were both classified into five groups, and a heatmap was generated utilizing the ComplexHeatmap package. Oncogenes and tumor suppressor genes (TSGs) were obtained from https://www.oncokb.org/cancer-genes. Venn diagrams were used to compare the overlap relationships between oncogenes, TSGs, pan-cancer prognostically favorable genes and prognostically unfavorable genes.

The rich annotation information from public databases was used to construct the knowledge network and enable a comprehensive and in-depth study of genes or drugs. The gene annotation data included GO (9), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (12), Reactome pathway (13), PathBank (14), chemical and genetic perturbations from MsigDB (15), drug perturbations from GEO (16), Disease Alliance (17), Clinvar disease (18), Human Phenotype Ontology (HPO) (19), Mondo Disease Ontology (MONDO) (20) and Library of Integrated Network-Based Cellular Signatures (LINCS) (21). Furthermore, drug targeting data were

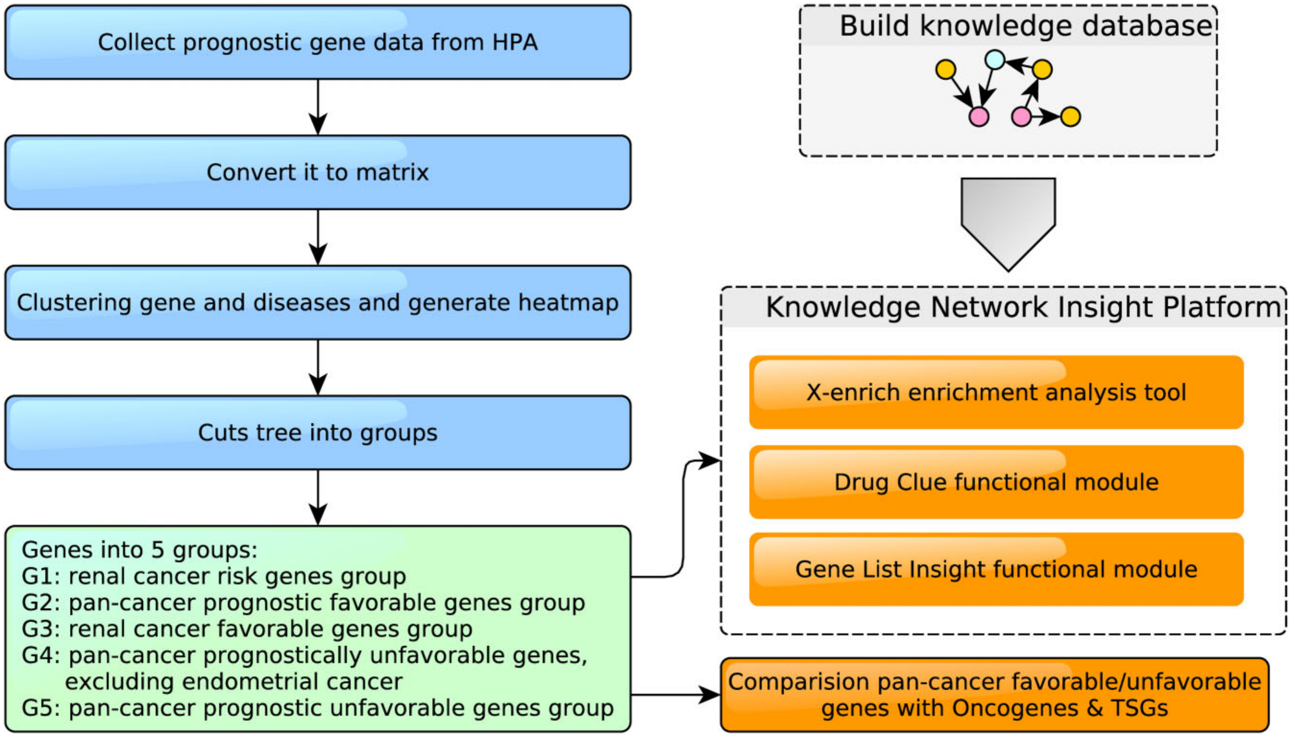

Collect prognostic gene data from HPA
Convert it to matrix
Clustering gene and diseases and generate heatmap
Cuts tree into groups
Genes into 5 groups:
G1: renal cancer risk genes group
G2: pan-cancer prognostic favorable genes group
G3: renal cancer favorable genes group
G4: pan-cancer prognostically unfavorable genes, excluding endometrial cancer
G5: pan-cancer prognostic unfavorable genes group
Build knowledge database
Knowledge Network Insight Platform
X-enrich enrichment analysis tool
Drug Clue functional module
Gene List Insight functional module
Comparision pan-cancer favorable/unfavorable genes with Oncogenes & TSGs

Figure 1. Overall process of this pan-cancer study.

obtained from DrugBank, and data on the effects of drugs on gene expression were obtained from CTD (22). Synthetic lethality gene pairs were obtained from SynLethDB (23), a comprehensive database that aggregates data from publications and four other related databases (Syn-lethality, Decipher, GenomeRNAi and BioGRID).

The GO was downloaded from http://geneontology.org/docs/download-ontology/. We extracted relationships such as ‘is_a’, ‘part_of’ and ‘regulates’ to construct relationships within the GO. GO terms were annotated to host proteins using the goa_human.gaf.gz file at http://current.geneontology.org/products/pages/downloads.html. The HPO was essential for understanding the relationship between phenotypes and genotypes. Ontology OBO and annotation HPOA files were obtained from https://hpo.jax.org/app/data/ and https://hpo.jax.org/app/data/annotations.

Data for drug–gene expression perturbation were downloaded from https://maayanlab.cloud/enrichr-kg/downloads (24). To ensure consistency, we map gene identifiers from different sources to Uniprot IDs. A total of 199 675 protein–protein interactions were extracted from the HINT (25) dataset (download from http://hint.yulab.org/download/), including direct and complex interactions. For the STRING database (26), we kept 839 224 high-confidence interactions with a score > 700. KEGG pathway information was converted into protein regulation network data by parsing the downloaded XML files.

By integrating the aforementioned data, we stored gene annotation information, protein–protein interactions, drug–gene interactions and phenotype–gene interactions in the Neo4j graph database, which enabled the visualization of intricate biomolecule interactions and flexible graph retrieval. The structure of this knowledge repository is depicted in Figure 2.

In our methodology, we utilize the R programming language to execute SQL statements and the Cypher query language for retrieving data stored within SQLite and Neo4J databases. To visualize biomolecular interaction networks, we use the vis.js library, enabling interactive network diagrams. This interactive visualization allows effective representation of nodes and edges, considering their attributes and relationships. Additionally, we use the plotly R package to generate interactive charts, improving the clarity and comprehensibility of our analysis results.

Given the intricate nature of the network, we employ centrality calculation methods to pinpoint pivotal nodes. Our platform incorporates four centrality measurements: degree, betweenness, closeness and eigenvector centrality. These measurements quantify the importance of nodes within the network, facilitating focused examinations of key biomolecules, drug targets or disease associations. The platform is accessible at https://medcode.link/cancerpro, and Figure 1 provides an overview of the system design.

We specifically designed three functional modules to analyze pan-cancer prognosis genes comprehensively to meet our research objectives. The first module, named X-enrich, focuses on functional enrichment analysis. We use the Over-Representation Analysis method to test whether known biological functions or processes are over-represented in a given gene list. Input genes are first mapped to the selected annotation sets. Subsequently, a hypergeometric test is conducted to identify over-represented terms within these sets, using all genes associated with the selected annotations as a reference. Since the annotation information not only encompasses functional annotations such as GO and pathways but also includes regulations of drugs, diseases, phenotypes, etc., we can analyze whether a gene list is enriched in these diverse entities. Based on the X-enrich module, we performed enrichment analysis of

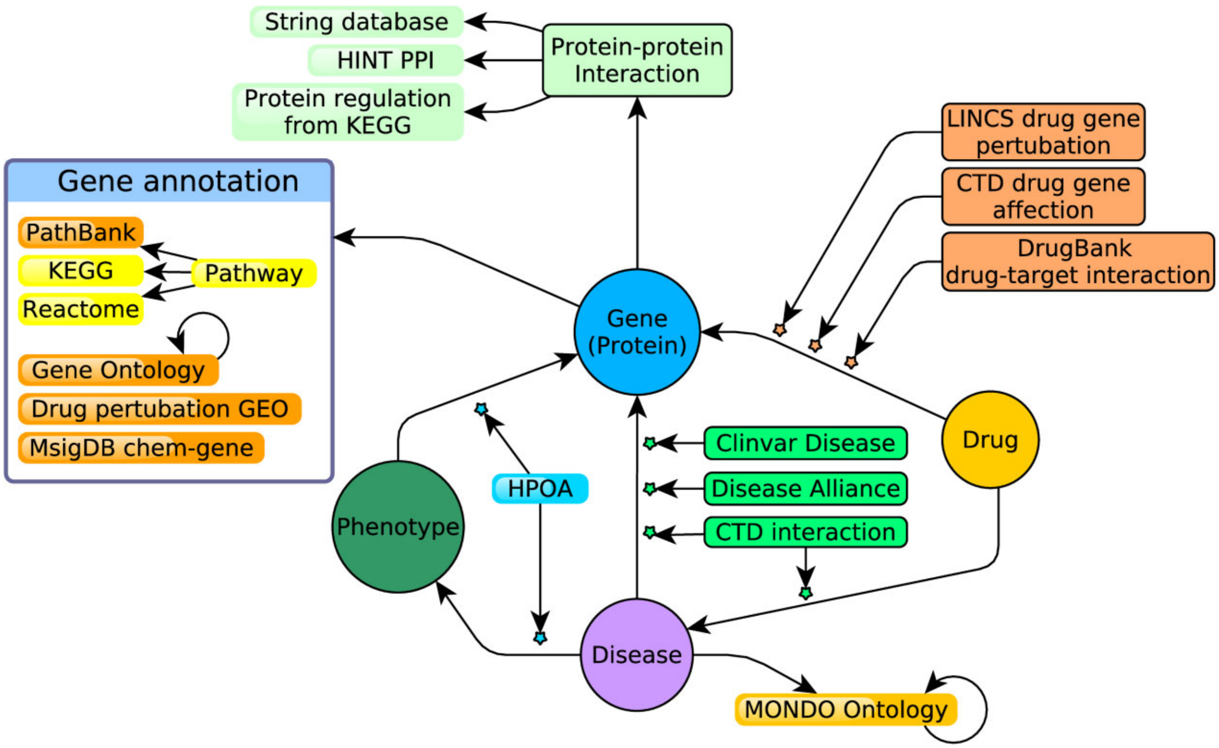


Figure 2. The organizational structure of the back-end knowledge repository. Asterisks on the continuous edges indicate the data sources corresponding to the respective relationships.

pan-cancer prognostically unfavorable genes based on KEGG pathway annotations. Additionally, utilizing LINCS annotation data, we employed enrichment analysis to identify drugs that significantly down-regulate these genes.

The second module focuses on a novel perspective on drugs called Drug Clue. By considering drug target proteins, genes affected by drugs and genes up-regulated or down-regulated by drugs, Drug Clue can identify the biological pathways, diseases and phenotypes associated with a particular drug. It can also identify drugs that may have similar effects. This information can be used to better understand how drugs work and to identify potential side effects. Using this Drug Clue module, we gain insight into drugs that significantly down-regulate pan-cancer prognostic unfavorable genes.

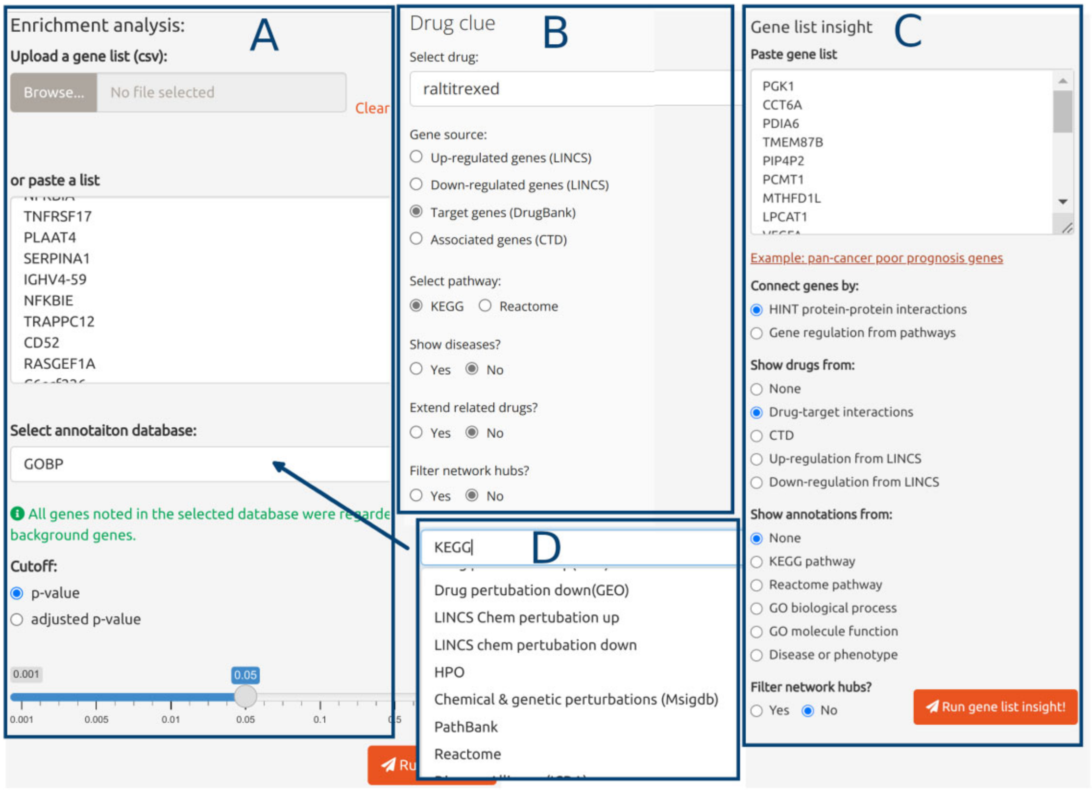
The third section introduces a new method for examining gene lists known as Gene List (GL) Insight. Through this module, genes are interconnected using highly reliable protein–protein interaction information or protein regulation information. Subsequently, we extracted drugs associated with two or more genes, either targeting these genes directly or affecting/regulating their expression. Additionally, the module can retrieve biological pathways from KEGG or Reactome, GO annotations and disease/phenotype information linked to multiple genes. Furthermore, the module offers support for selecting critical nodes based on the four centrality measurements mentioned above. We conducted a detailed analysis of prognostically unfavorable genes specific to prostate cancer in this module. Using HINT, we connected them and retrieved Reactome pathway information. In our study, we identified 43 genes that are prognostically unfavorable in at least five different types of cancer. Using the GL Insight module, we connected them with high-confidence protein–protein interactions from HINT and obtained GO biological process annotations associated with at least two genes.

Results

Knowledge network insight platform construction

We have constructed a knowledge network that encompasses a broad range of biomedical information. Gene annotations include details from pathways, GO and drug perturbations. Interaction information encompasses protein–protein interaction, gene–phenotype relationships, disease–phenotype relationships, disease–disease relationships, drug–gene regulation data, drug–target interaction and gene–disease association information. Unlike typical gene annotations, ours extend beyond functional annotations to include disease phenotypes, drug interference and targeted drug information. This allows us to classify genes based on multiple features and conduct enrichment analyses.

The CancerPro knowledge insight platform was developed to provide three key functions: X-enrich enrichment analysis based on various types of annotation information, Drug Clue and GL Insight. These functions are illustrated in Figure 3. Figure 3A depicts the X-enrich enrichment analysis module. Hypergeometric testing is employed to assess the significant associations between input gene lists and annotations such as GO, KEGG, Reactome, drug perturbation, phenotype and disease. Figure 3B illustrates the Drug Clue function, enabling the retrieval of genes targeted by a drug. Furthermore, it can identify genes influenced by a drug derived from the LINCS dataset. By analyzing associations of genes, we can find similar drugs that target these genes, and we can also examine the related diseases and phenotypes associated with them. The module also offers visualizations of the biological pathways associated with these genes, allowing for a multidimensional study of drug characteristics. Through the Drug Clue user interface, users can obtain specific clue information on 7437 different drugs based on their selected criteria.


Enrichment analysis:

Upload a gene list (csv):

No file selected

or paste a list

TNFRSF17
PLAAT4
SERPINA1
IGHV4-59
NFKBIE
TRAPPC12
CD52
RASGEF1A

Select annotaiton database: GOBP

All genes noted in the selected database were regarde background genes.

Cutoff:
● p-value
○ adjusted p-value

Drug clue

Select drug: raltitrexed

Gene source:
○ Up-regulated genes (LINCS)
○ Down-regulated genes (LINCS)
● Target genes (DrugBank)
○ Associated genes (CTD)

Select pathway: ● KEGG ○ Reactome

Show diseases? ○ Yes ● No

Extend related drugs? ○ Yes ● No

Filter network hubs? ○ Yes ● No

Gene list insight

Paste gene list

PGK1
CCT6A
PDIA6
TMEM87B
PIP4P2
PCMT1
MTHFD1L
LPCAT1

Example: pan-cancer poor prognosis genes

Connect genes by:
● HINT protein-protein interactions
○ Gene regulation from pathways

Show drugs from:
○ None
● Drug-target interactions
○ CTD
○ Up-regulation from LINCS
○ Down-regulation from LINCS

Show annotations from:
● None
○ KEGG pathway
○ Reactome pathway
○ GO biological process
○ GO molecule function
○ Disease or phenotype

Filter network hubs? ○ Yes ● No

KEGG

Drug pertubation down(GEO)
LINCS Chem pertubation up
LINCS chem pertubation down
HPO
Chemical & genetic perturbations (Msigdb)
PathBank
Reactome

Figure 3. Pan-cancer insight platform functional module user interface. (**A**) The X-enrich enrichment analysis module. (**B**) The Drug Clue function. (**C**) The Gene List Insight module. (**D**) The gene annotation sources list.

Figure 3C presents the GL Insight module, which connects input gene lists through HINT protein–protein interactions or protein regulation information. It analyzes the inherent interactions within the gene list, providing information on drugs associated with the at least two input genes. These associations can be either targeted relationships or relationships involving gene regulation or interference. This module also extracts biological functional annotations, such as GO, KEGG, Reactome, diseases and phenotypes, that are related to two or more genes. This offers a novel perspective on observing gene lists.

CancerPro incorporates additional features to enhance its functionality. Users can filter cancer-specific marker genes, enabling a more focused perspective. Interactive visualization tools, such as heatmaps, facilitate effortless exploration and comparison of pan-cancer prognostic gene expression patterns. Moreover, CancerPro enables interactive analysis of synthetic lethal gene pairs, empowering researchers to explore potential therapeutic strategies by identifying gene combinations where inhibiting one gene can be lethal to cancer cells carrying a mutation in another. In summary, CancerPro offers a comprehensive suite of tools for thorough analysis and exploration of drugs, genes and gene lists from various perspectives.

Hierarchical clustering analysis of prognostic marker genes in diverse cancer types

Hierarchical clustering was performed on both samples and genes using prognostic favorable/unfavorable gene labels. The result is depicted in Figure 4A, with detailed gene information shown in Supplementary Figure S1 and accessible through the CancerPro web server. The number of marker genes for each tumor is listed in Supplementary Table S1, sheet marker_gene_number. Using cluster analysis, tumors were classified into five prognostic groups, denoted as P1–P5. P1 is dominated by endometrial cancer, while P2 is the most diverse group, encompassing various cancers such as cervical, head and neck, urothelial, breast, melanoma, stomach, colorectal, ovarian, thyroid, glioma, prostate and testis cancers. Interestingly, prostate and testis cancers share similarities, as do cervical and head and neck cancers. Notably, P2 is enriched with digestive system cancers. The prostate and testis cancers are characterized by fewer prognostic markers, and are called ‘pure cancers’. Glioma and thyroid tumors also have relatively few marker genes. P3 comprises renal cancer, exhibiting a prognostic gene profile distinct from the others. As shown in Figure 4A, genes at the top are associated with poor prognosis in renal cancer but good prognosis in other diseases. The P4 group is characterized by liver cancer, while the P5 group encompasses lung and pancreatic cancer.

Gene clustering analysis resulted in five gene groups, labeled G1–G5. G1 appears to be a renal cancer risk gene group. Most genes in this group are associated with unfavorable prognosis in renal cancer and favorable prognosis in other tumors. G2 represents the pan-cancer prognostic favorable genes group, while G3 comprises genes associated with good outcomes in renal cancer but indicates poor prognosis in other tumors. G4 is a group of pan-cancer prognostic unfavorable genes, excluding cervical cancer. Finally, G5 forms the pan-cancer prognostic unfavorable gene group.

This clustering and classification analysis provide valuable insights into the diverse prognostic marker landscapes across different cancer types.

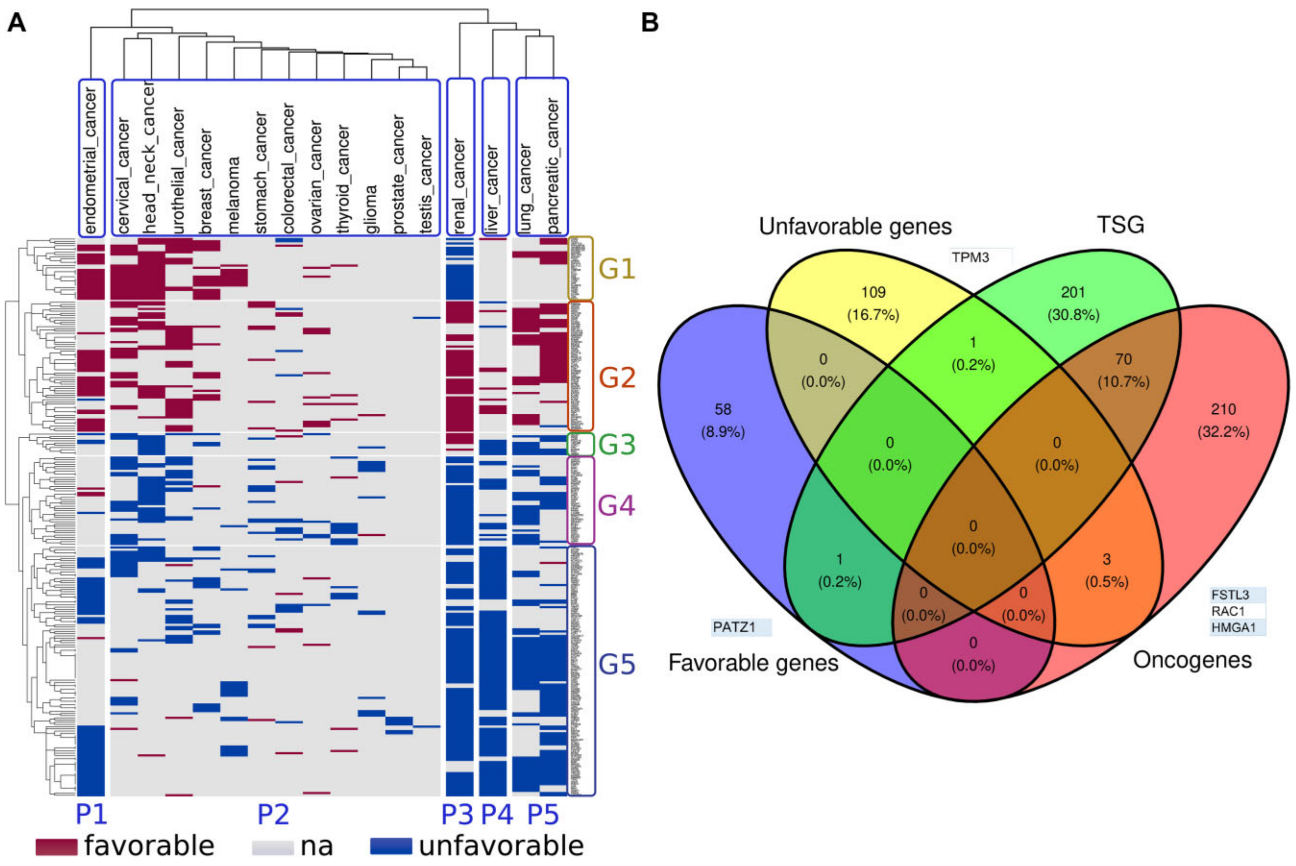


Figure 4. Pan-cancer clustering based on marker gene status. (**A**) Tumors were classified into five prognostic groups, denoted as P1–P5. Gene clustering analysis resulted in five gene groups, labeled G1–G5. The interactive full-size figure is accessible through the online CancerPro server. (**B**) Venn diagram of oncogenes, tumor suppressor genes, pan-cancer prognostic favorable genes and pan-cancer prognostic unfavorable genes.

Limited overlap between pan-cancer prognostic genes and oncogenes/TSGs

In Figure 4B, the Venn diagram shows the number of oncogenes (283), TSGs (173) and both pan-cancer prognostic genes (G2) and unfavorable genes (G5). The figure shows that nearly all cancer prognostic genes do not overlap with TSGs or oncogenes. Only three pan-cancer prognostic unfavorable genes (FSTL3, RAC1 and HMGA1) are also oncogenes. Only one gene, PATZ1, is both a tumor suppressor gene and a pan-cancer prognostic favorable gene.

Analysis of pan-cancer prognostic unfavorable genes

Supplementary Table S1, sheet 'Top_unfavorable_gene,' presents a list of 43 prognostic unfavorable genes, associated with at least five different cancer types. We utilized the GL Insight functional module, using high-confidence protein–protein interactions to connect these genes. As illustrated in Figure 5A, this formed a local cluster with high connectivity.

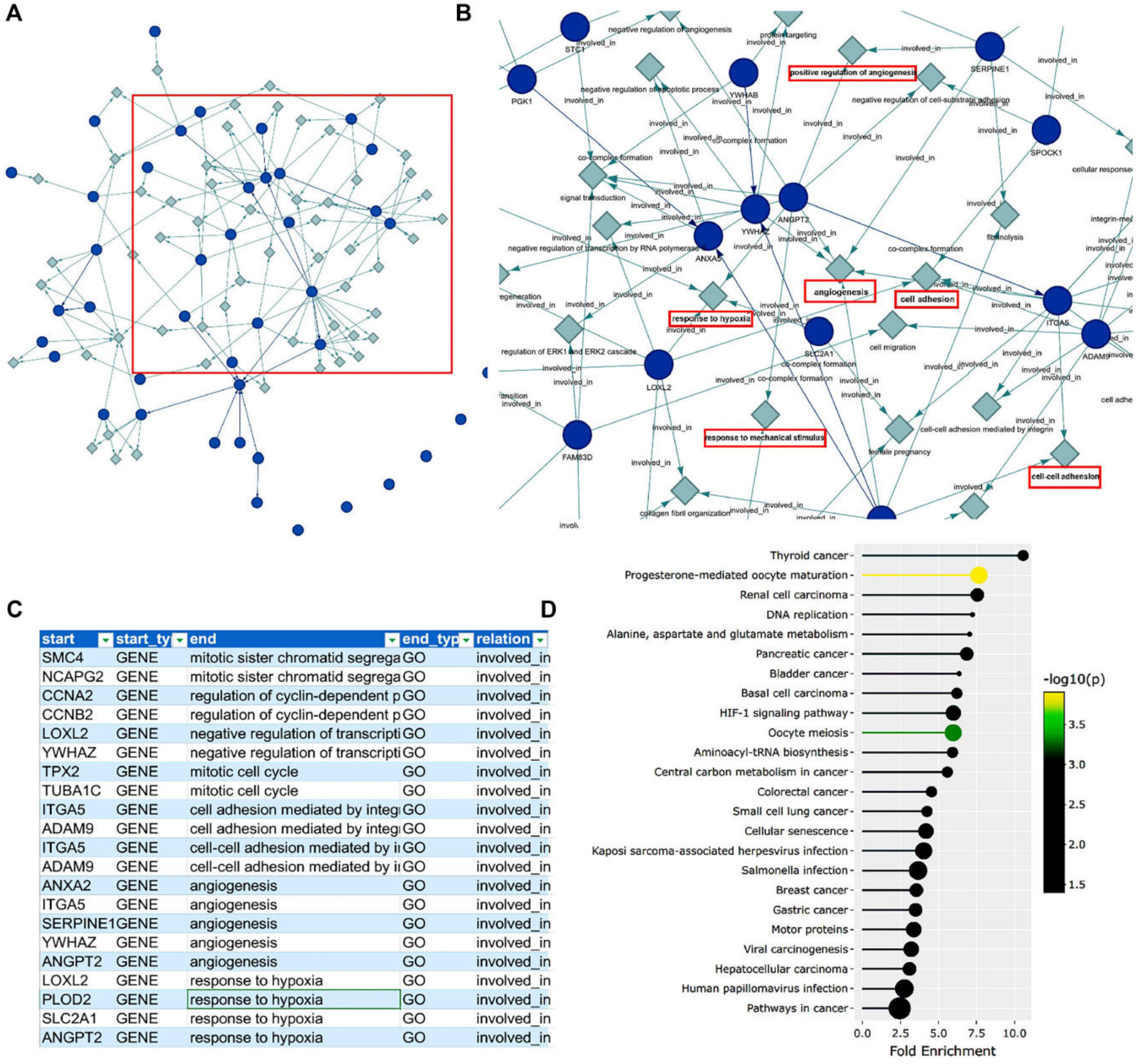
Further examination of the network revealed the involvement of these genes in crucial biological processes such as angiogenesis, cell adhesion and hypoxia response (Figure 5B, C). Details of all interactions between these entities can be found in Supplementary Table S2, sheet 'Top_gene_relations'.

Using X-enrich and the hypergeometric test algorithm, we conducted a KEGG functional enrichment analysis of gene cluster G5, which contains pan-cancer prognostic unfavorable genes, as shown in Figure 5D. The analysis revealed that these genes are involved in diverse cancer pathways, including pancreatic cancer, colorectal cancer, small cell lung cancer and breast cancer. Additionally, they are implicated in biological pathways such as DNA replication, viral carcinogenesis, human papillomavirus infection and potentially others.

There are seven genes involved in progesterone-mediated oocyte maturation and oocyte meiosis pathways, namely CDK1, CCNB2, CCNA2, BUB1, AURKA, PLK1 and YWHAZ. Of these, CCNA2, BUB1, AURKA and YWHAZ are specific high-expression marker genes for endometrial cancer.

A GL Insight analysis of gene cluster G5 identified drugs that target at least two or more proteins. Using a degree-based algorithm, one of the important drugs was found to be copper, which targets LDHA and PGK1, both of which are ligands of Artenimol. The growth and spread of cancer cells in the human body depend on proteins that bind to copper ions. Research on how cancer-related proteins interact with metals and with other proteins provides clues for potential new cancer drug targets (27).

Phenethyl isothiocyanate (PEITC) targets TPM3, S100A10 and YWHAZ. Epidemiological evidence suggests that there is a strong inverse relationship between cruciferous vegetable intake and cancer incidence. PEITC is naturally found in cruciferous vegetables such as mustard, cabbage and broccoli. It is well known that PEITC targets multiple proteins to inhibit a variety of cancer-promoting mechanisms, such as cell proliferation, progression and metastasis. Additionally, preclinical evidence suggests that it is very effective when combined with conventional anti-cancer drugs to improve overall



start	start_ty	end	end_typ	relation
SMC4	GENE	mitotic sister chromatid segrega	GO	involved_in
NCAPG2	GENE	mitotic sister chromatid segrega	GO	involved_in
CCNA2	GENE	regulation of cyclin-dependent p	GO	involved_in
CCNB2	GENE	regulation of cyclin-dependent p	GO	involved_in
LOXL2	GENE	negative regulation of transcripti	GO	involved_in
YWHAZ	GENE	negative regulation of transcripti	GO	involved_in
TPX2	GENE	mitotic cell cycle	GO	involved_in
TUBA1C	GENE	mitotic cell cycle	GO	involved_in
ITGA5	GENE	cell adhesion mediated by integ	GO	involved_in
ADAM9	GENE	cell adhesion mediated by integ	GO	involved_in
ITGA5	GENE	cell-cell adhesion mediated by i	GO	involved_in
ADAM9	GENE	cell-cell adhesion mediated by i	GO	involved_in
ANXA2	GENE	angiogenesis	GO	involved_in
ITGA5	GENE	angiogenesis	GO	involved_in
SERPINE1	GENE	angiogenesis	GO	involved_in
YWHAZ	GENE	angiogenesis	GO	involved_in
ANGPT2	GENE	angiogenesis	GO	involved_in
LOXL2	GENE	response to hypoxia	GO	involved_in
PLOD2	GENE	response to hypoxia	GO	involved_in
SLC2A1	GENE	response to hypoxia	GO	involved_in
ANGPT2	GENE	response to hypoxia	GO	involved_in

Figure 5. Analysis of pan-cancer prognostic unfavorable genes. (**A**) These genes form a cluster with high connectivity. (**B**) Biological processes involved with these genes. (**C**) The downloadable network result from (**B**). (**D**) Functional enrichment analysis uses X-enrich module for these genes.

efficacy. PEITC appears to be a promising cancer treatment drug and has already been in clinical trials for leukemia and lung cancer (28).

Using the X-enrich function, we selected the 'LINCS chem perturbation down' annotation to analyze which drugs significantly down-regulate pan-cancer prognostic unfavorable genes. The analysis results are shown in Supplementary Table S2, sheet Chem_perturbation_down. We found that Raltitrexed, panobinostat, 9,10-deepithio-9,10-dehydrocanthifolicin, genistein and hycanthone significantly down-regulated poor prognosis genes in the G5 group.

In-depth investigation into the drug Raltitrexed by Drug Clue module

We configured the parameters with 'Gene source' set to Target genes (DrugBank), selected Reactome as the pathway, 'Show diseases' as Yes and excluded the extension of related drugs. The analysis revealed that Raltitrexed targets TYMS, and the drug acts as an inhibitor of TYMS. Figure 6 illustrates the Reactome pathways associated with this protein and provides information on related diseases.

As shown in Figure 6, the target protein TYMS is implicated in acute lymphoblastic leukemia, hepatocellular carcinoma, acute myeloid leukemia, rheumatoid arthritis and plasma cell myeloma. TYMS is a marker for colorectal adenocarcinoma via orthology data, and is involved in Reactome pathways including interconversion of nucleotide di- and triphosphates, and G_1/S-specific transcription.

Pan-cancer prognostic favorable genes G2

Upon utilizing the GL Insight for this gene set and employing protein connectivity through the HINT database, along with

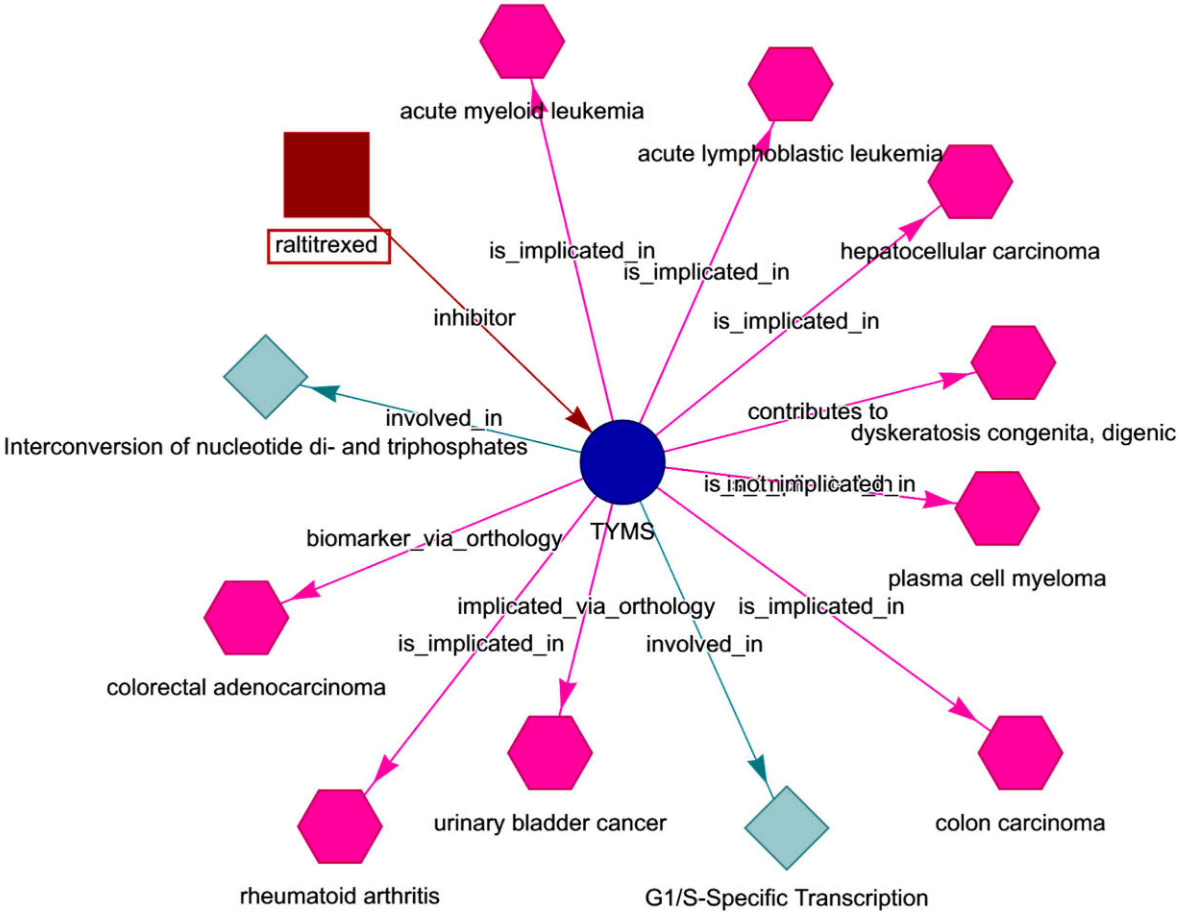


Figure 6. Investigation of the drug Raltitrexed by the Drug Clue module.

KEGG pathway annotation, we discovered that the associated KEGG pathways, including metabolic pathways, carbon metabolism, fatty acid metabolism/degradation and valine–leucine–isoleucine degradation, may play pivotal roles.

Using drug–target information to connect these genes, we identified glutathione (GSH) targeting GPX4 and HAGH, and flavin adenine dinucleotide (FAD) targeting ACAD8 and ACADS. Recent studies have emphasized the significance of GSH in key signal transduction reactions, acting as a controller of cell differentiation, proliferation, apoptosis, ferroptosis and immune function. Molecular alterations in the GSH antioxidant system and disturbances in GSH homeostasis have been implicated in tumor initiation, progression and treatment response. FAD acts as a biosensor of metabolic states, indicating a potential connection between epigenetics and metabolism, particularly in cancer cells. A study has observed that up-regulating *S*-adenosyl-L-methionine (SAM) and FAD production, while down-regulating acetyl-CoA, NAD and tetrahydrofolate (THF), could be reasonable targets for inhibiting tumor cells (29).

Connecting these genes through diseases, we discovered that obesity is linked to the GNG7 and PLIN5 genes. While overweight and obesity are generally considered to increase the risk of diseases and mortality, there is a phenomenon known as the ‘obesity paradox’, where overweight and obesity may show an association with lower mortality risk for specific diseases. Large-scale studies across various cancer types examining the relationship between body weight and survival suggest that a higher body mass index (BMI) during the diagnostic period, especially in cases of overweight or mild obesity, is associated with improved cancer patient survival rates (30).

Analysis of renal and prostate cancer risk genes

We investigated the G1 group of genes associated with renal cancer risk. Most of these genes are poor prognostic markers in renal cancer and prognostically favorable genes in other cancers. First, we used X-enrich enrichment analysis to study diseases associated with mutations in these genes. We selected the annotations from the ClinVar disease database, and the results are shown in Supplementary Table S2, sheet Renal_risk_enrichment.

The results showed that most of these genes are related to immunodeficiency, including TYK2, CD27, TRAC, CD3E and CD3D. In renal cancer, high expression of these genes is an unfavorable factor, but in other cancers they are protective factors. The possible reason is that immunodeficiency impacts cancer prognosis depending on the type of cancer. Some cancers may be more susceptible to immune system control, while others may be less influenced by immune function.

We extracted 51 poor prognostic genes that are specifically present in prostate cancer but not in other cancers, as shown in Supplementary Table S1, sheet ‘Prostate_specific_markers’. We used the GL Insight module to analyze this gene list, utilizing HINT to connect proteins and Reactome annotations to annotate more than two genes. The results are shown in Supplementary Figure S2. We found that abnormal alternative splicing significantly impacts the characteristics of prostate cancer cells.

Discussion

We provide an innovative perspective on the study of pan-cancer prognostic genes, forming a complete landscape of the relationship between gene markers and tumor prognosis. Our development, CancerPro, is a versatile knowledge network platform that is not only limited to pan-cancer prognostic gene analysis. By integrating extensive data on genes, drugs, diseases and their interactions, the platform provides flexible tools for investigating a wide range of biological questions. Researchers can explore other interesting findings, such as genes, gene lists and drugs, through our CancerPro platform, expanding research possibilities.

Through clustering analysis, we found that prostate and testis cancers have very few prognostic marker genes. Other tumors with relatively few marker genes include glioma and thyroid, which we refer to as ‘pure tumors’. These tumors may be associated with specific carcinogenic mechanisms significantly different from other tumors. Prostate cancer has fewer prognostic marker genes than other cancers. It also exhibits a fairly low mutation burden, meaning that the prostate cancer genome is quite stable. The prostate is part of the male reproductive system, and may have a relatively high DNA repair capacity. The low prostate cell proliferation rate may reduce mutation opportunities. Prostate cancer incidence may be associated with relatively little environmental exposure, and thus the opportunity for exposure to DNA damage may be lower. Mutation burden is an independent prognostic factor for pan-cancer survival, but the relationship is nonlinear (31). For prostate cancer, studies have shown that patients in the high-TMB group had lower overall survival than those in the low-TMB group (32). These findings demonstrate prostate cancer’s uniqueness. Our research found that alternative splicing plays a significant role in prostate cancer development. Using pathway-guided analysis, Phillips *et al.* identified Myc-dependent alternative pre-mRNA splicing in aggressive prostate cancers (33). In order to identify genes that regulate alternative splicing in cancer, the researchers mined a large amount of prostate cancer.

Our study found that there is very limited overlap between prognosis marker genes and oncogenes and TSGs. This suggests that they play different roles in tumor development. Oncogenes and TSGs may play a more critical role in the early stages of tumor development. Mutations or abnormal expression of oncogenes may be a key factor driving tumor development. The enhanced activity of these genes may promote excessive cell proliferation and escape normal growth regulatory mechanisms. Mutations or inactivations of TSGs may cause cells to lose their normal inhibitory effect on cancer, promoting abnormal cell proliferation. In the middle and late stages of the disease, oncogenes and TSGs may only maintain proliferation and survival of tumor cells, but some targeted therapy strategies can inhibit these genes. Prognostic marker genes may play a more significant role in the middle and late stages of tumor development. They are closely related to disease progression and patient survival. They may be used to determine the tumor grade and stage, providing detailed information about the tumor, including its biological characteristics and prognosis. These markers can predict patient survival. This information could be utilized to develop new targeted therapies for cancer.

Renal cancer is clearly distinguished from other tumors in the clustering analysis, indicating that its driving factors may be different from those of other tumors. For example, renal cancer may be significantly associated with genetic factors (34). Our study found that the renal cancer-specific genome is associated with immunodeficiency. These genes include TYK2, CD27, TRAC, CD3E and CD3D. It is not favorable if these genes are highly expressed in renal cancer. However, they may be protective factors for other cancers, such as endometrial cancer, cervical cancer, head and neck cancer, urothelial cancer, breast cancer and melanoma. This suggests that the immune system may respond differently to different cancer types. In renal cancer, it suppresses the immune system, but in other cancers, it activates immune cells. These genes are probably key potential targets for renal cancer treatment.

In the KEGG enrichment analysis of pan-cancer prognostically unfavorable genes G5, seven genes involved in the two biological pathways progesterone-mediated oocyte maturation and oocyte meiosis, namely CDK1, CCNB2, CCNA2, BUB1, AURKA, PLK1 and YWHAZ, of which CCNA2, BUB1, AURKA and YWHAZ, are endometrial cancer-specific high-expression markers. CCNA2 encodes a protein that belongs to the highly conserved family of cyclins, which regulate the cell cycle by promoting transitions from G_1/S. The BUB1 gene encodes a serine/threonine protein kinase that plays a central role in mitosis. BUB1β-binding kinase showed a consistent association with improved overall survival (35). The AURKA protein plays a crucial role in forming and stabilizing microtubules, which are structures essential for separating chromosomes during cell division. A study has shown that increased levels of AURKA can promote the uncontrolled growth of cells (proliferation) and are associated with a poorer prognosis in bladder cancer patients (36). YWHAZ is involved in the pathway of activation of BAD (BCL2-associated agonist of cell death) and translocation to mitochondria. Prostate cancer patients with high levels of YWHAZ expression were found to have an increased risk of developing castration-resistant prostate cancer and a shorter overall survival time (37).

Enrichment analysis of pan-cancer prognostically unfavorable genes G5 using the X-enrich function revealed that Raltitrexed, Panobinostat, 9,10-deepithio-9,10-dehydrocanthifolicin, Genistein and Hycanthone significantly down-regulated these genes. Raltitrexed is an anti-neoplastic agent and folic acid antagonist. Raltitrexed inhibits thymidylate synthase, leading to DNA fragmentation and cell death. Panobinostat is a histone deacetylase inhibitor (38). It acts through multiple pathways, including regulating the cell cycle, inducing apoptosis and suppressing angiogenesis, by affecting chromatin structure and gene expression. The drug 9,10-deepithio-9,10-dehydrocanthifolicin, also known as okadaic acid, inhibits phosphoserine/threonine protein phosphatases 1 and 2a. It is also a potent tumor promoter. Genistein is a plant-derived compound belonging to the isoflavone family. It is mainly found in soybeans and other legumes. It has a variety of biological activities, including antioxidant, anti-inflammatory, anti-cancer and hormone-regulating effects (39). Our analysis showed its potential inhibitory effects on various tumors. Hycanthone’s metabolites generate oxygen radicals, which are highly reactive molecules that can harm cell structure and function (40). This may be one mechanism by which hycanthone causes cell damage and death. The above analysis suggests that these drugs may have broad anti-tumor effects.

A novel analysis of Raltitrexed using the Drug Clue module revealed that its target protein is TYMS. TYMS is implicated in acute lymphoblastic leukemia, hepatocellular carcinoma, acute myeloid leukemia, rheumatoid arthritis and plasma cell myeloma. It is a marker for colorectal adenocarcinoma. It is involved in Reactome pathways including interconversion of nucleotide di- and triphosphates, and G_1/S-specific transcription. The interconversion of nucleotide di- and triphosphates plays a crucial role in various cellular processes, including DNA replication, RNA synthesis and energy metabolism. These processes are essential for cell growth and proliferation and can be linked to cancer in several ways. Cell cycle regulation and cancer development are closely linked to G_1/S-specific transcription. DNA damage must be repaired at the $G_{1/}$S transition in order for the cell to progress to DNA synthesis in the S phase. Uncontrolled cell proliferation can be caused by dysregulation of G_1/S-specific transcription. We would like to point out that this demonstrates the powerful capabilities of the Drug Clue module in terms of knowledge presentation and knowledge discovery.

Our research is a significant contribution to the study of pan-cancer prognostic marker genes, providing a comprehensive and novel perspective on this field of research. Furthermore, we provide a flexible fresh CancerPro Knowledge Network Insight platform that can be used to study other biomedical problems. The current study is limited by its focus on gene expression-based prognostic markers. To address this, future research could incorporate multi-omics datasets and develop specialized analytical tools. Shared cancer genes may act differently due to tissue and molecular differences, highlighting the need to focus on tumor heterogeneity and precision medicine. The platform's unique feature lies in its ability to present knowledge to researchers in a novel and powerful manner, enabling them to uncover previously overlooked connections and spark new research ideas. Knowledge graphs are often dynamic and incomplete, making link prediction a valuable method for filling in missing information. By doing so, we can uncover potential associations for research, such as predicting novel (potentially prognostic) genes by examining the properties of known prognostic markers. However, this process is resource intensive, which may make it unsuitable for integration into CancerPro, or may require optimization such as pre-computing and storing results in the backend to improve usability.

Data availability

The code for processing raw HPA data and clustering analysis can be freely obtained from https://github.com/Cetomato/CancerPro (DOI: https://doi.org/10.5281/zenodo.14010461). The Drug Clue, X-enrich (enrichment analysis) and Gene List (GL) Insight functional modules are accessible from the https://medcode.link/cancerpro main menu. All analysis results can be downloaded from the dedicated analysis results download panel.

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

We appreciate the suggestions provided in this study by Wei Liu, Department of Anatomy and Histology, and Pencheng Shu, State Key Laboratory of Medical Molecular Biology, Department of Molecular Biology and Biochemistry. We also thank Zunfan Chen, Zuda Li, and Chenjun Zheng from the Department of Biomedical Engineering at PUMC for their help. Special thanks to Lifang Xie from the Information Technology Center of PUMC, as well as Furen Shi and Sheng Yang from the Information Center of IBMS for their support.

Funding

CAMS Innovation Fund for Medical Sciences (No. 2021-I2M-1-057); and Teaching Reform Project, School of Basic Medicine, Peking Union Medical College (No. 2023jcjg0107).

Conflict of interest statement

None declared.

References

1. Tomczak,K., Czerwińska,P. and Wiznerowicz,M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn)*, **19**, A68–A77.
2. Cancer,Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**, 543–550.
3. Pusztai,L., Mazouni,C., Anderson,K., Wu,Y. and Symmans,W.F. (2006) Molecular classification of breast cancer: limitations and potential. *Oncologist*, **11**, 868–877.
4. Chen,B., Khodadoust,M.S., Liu,C.L., Newman,A.M. and Alizadeh,A.A. (2018) Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.*, **1711**, 243–259.
5. Shen,L., Shi,Q. and Wang,W. (2018) Double agents: genes with both oncogenic and tumor-suppressor functions. *Oncogenesis*, **7**, 25.
6. Han,S., Kim,H.Y., Park,K., Cho,H.J., Lee,M.S., Kim,H.J. and Kim,Y.D. (1999) c-myc expression is related with cell proliferation and associated with poor clinical outcome in human gastric cancer. *J. Korean Med. Sci.*, **14**, 526–530.
7. Ji,F., Zhang,Z.-H., Zhang,Y., Shen,S.-L., Cao,Q.-H., Zhang,L.-J., Li,S.-Q., Peng,B.-G., Liang,L.-J. and Hua,Y.-P. (2018) Low expression of c-myc protein predicts poor outcomes in patients with hepatocellular carcinoma after resection. *BMC Cancer*, **18**, 460.
8. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
9. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*, **25**, 25–29.
10. Slenter,D.N., Kutmon,M., Hanspers,K., Riutta,A., Windsor,J., Nunes,N., Mélius,J., Cirillo,E., Coort,S.L., Digles,D., *et al.* (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, **46**, D661–D667.
11. Thul,P.J. and Lindskog,C. (2018) The human protein atlas: a spatial map of the human proteome. *Protein Sci.*, **27**, 233–244.
12. Kanehisa,M., Furumichi,M., Sato,Y., Kawashima,M. and Ishiguro-Watanabe,M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.
13. Fabregat,A., Korninger,F., Viteri,G., Sidiropoulos,K., Marin-Garcia,P., Ping,P., Wu,G., Stein,L., D'Eustachio,P. and

Hermjakob,H. (2018) Reactome graph database: efficient access to complex pathway data. *PLoS Comput. Biol.*, **14**, e1005968.

14. Wishart,D.S., Li,C., Marcu,A., Badran,H., Pon,A., Budinski,Z., Patron,J., Lipton,D., Cao,X., Oler,E., *et al.* (2020) PathBank: a comprehensive pathway database for model organisms. *Nucleic Acids Res.*, **48**, D470–D478.
15. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdóttir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.
16. 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., *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
17. Collins,R., Balaconis,M.K., Brunak,S., Chen,Z., Silva,M.D., Gaziano,J.M., Ginsburg,G.S., Jha,P., Kuri,P., Metspalu,A., *et al.* (2022) Global priorities for large-scale biomarker-based prospective cohorts. *Cell Genomics*, **2**, 100141.
18. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W., *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
19. Köhler,S., Gargano,M., Matentzoglu,N., Carmody,L.C., Lewis-Smith,D., Vasilevsky,N.A., Danis,D., Balagura,G., Baynam,G., Brower,A.M., *et al.* (2020) The human phenotype ontology in 2021. *Nucleic Acids Res.*, **49**, D1207–D1217.
20. Shefchek,K.A., Harris,N.L., Gargano,M., Matentzoglu,N., Unni,D., Brush,M., Keith,D., Conlin,T., Vasilevsky,N., Zhang,X.A., *et al.* (2022) The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.*, **48**, D704–D715.
21. Musa,A., Tripathi,S., Dehmer,M. and Emmert-Streib,F. (2019) L1000 Viewer: a search engine and web interface for the LINCS data repository. *Front. Genet.*, **10**, 557.
22. Davis,A.P., Grondin,C.J., Johnson,R.J., Sciaky,D., Wiegers,J., Wiegers,T.C. and Mattingly,C.J. (2021) Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Res.*, **49**, D1138–D1143.
23. Wang,J., Wu,M., Huang,X., Wang,L., Zhang,S., Liu,H. and Zheng,J. (2022) SynLethDB 2.0: a web-based knowledge graph database on synthetic lethality for novel anticancer drug discovery. *Database (Oxford)* **2022**, baac030.
24. Evangelista,J.E., Clarke,D.J.B., Xie,Z., Lachmann,A., Jeon,M., Chen,K., Jagodnik,K.M., Jenkins,S.L., Kuleshov,M.V., Wojciechowicz,M.L., *et al.* (2022) SigCom LINCS: data and metadata search engine for a million gene expression signatures. *Nucleic Acids Res.*, **50**, W697–W709.
25. Das,J. and Yu,H. (2012) HINT: high-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.*, **6**, 92.
26. Szklarczyk,D., Franceschini,A., Wyder,S., Forslund,K., Heller,D., Huerta-Cepas,J., Simonovic,M., Roth,A., Santos,A., Tsafou,K.P., *et al.* (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.
27. Zhang,X., Walke,G.R., Horvath,I., Kumar,R., Blockhuys,S., Holgersson,S., Walton,P.H. and Wittung-Stafshede,P. (2022) Memo1 binds reduced copper ions, interacts with copper chaperone Atox1, and protects against copper-mediated redox activity in vitro. *Proc. Natl Acad. Sci. USA*, **119**, e2206905119.
28. Gupta,P., Wright,S.E., Kim,S.-H. and Srivastava,S.K. (2014) Phenethyl isothiocyanate: a comprehensive review of anti-cancer mechanisms. *Biochim. Biophys. Acta*, **1846**, 405–424.
29. He,Y., Gao,M., Tang,H., Cao,Y., Liu,S. and Tao,Y. (2019) Metabolic intermediates in tumorigenesis and progression. *Int. J. Biol. Sci.*, **15**, 1187–1199.
30. Tu,H., McQuade,J.L., Davies,M.A., Huang,M., Xie,K., Ye,Y., Chow,W.-H., Rodriguez,A. and Wu,X. (2022) Body mass index and survival after cancer diagnosis: a pan-cancer cohort study of 114 430 patients with cancer. *Innovation (Camb.)*, **3**, 100344.
31. Smith,J.R., Parl,F.F. and Dupont,W.D. (2023) Mutation burden independently predicts survival in the pan-cancer atlas. *JCO Precis. Oncol.*, **7**, e2200571.
32. Wang,L., Pan,S., Zhu,B., Yu,Z. and Wang,W. (2021) Comprehensive analysis of tumour mutational burden and its clinical significance in prostate cancer. *BMC Urol.*, **21**, 29.
33. Phillips,J.W., Pan,Y., Tsai,B.L., Xie,Z., Demirdjian,L., Xiao,W., Yang,H.T., Zhang,Y., Lin,C.H., Cheng,D., *et al.* (2020) Pathway-guided analysis identifies myc-dependent alternative pre-mRNA splicing in aggressive prostate cancers. *Proc. Natl Acad. Sci. USA*, **117**, 5269–5279.
34. Nabi,S., Kessler,E.R., Bernard,B., Flaig,T.W. and Lam,E.T. (2018) Renal cell carcinoma: a review of biology and pathophysiology. *F1000Res*, **7**, 307.
35. Ocaña,A., Pérez-Peña,J., Díez-González,L., Sánchez-Corrales,V., Templeton,A., Seruga,B., Amir,E. and Pandiella,A. (2016) Transcriptomic analyses identify association between mitotic kinases, PDZ-binding kinase and BUB1, and clinical outcome in breast cancer. *Breast Cancer Res. Treat.*, **156**, 1–8.
36. Guo,M., Lu,S., Huang,H., Wang,Y., Yang,M.Q., Yang,Y., Fan,Z., Jiang,B. and Deng,Y. (2018) Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. *BMC Syst. Biol.*, **12**, 118.
37. Rüenauver,K., Menon,R., Svensson,M.A., Carlsson,J., Vogel,W., Andrén,O., Nowak,M. and Perner,S. (2014) Prognostic significance of YWHAZ expression in localized prostate cancer. *Prostate Cancer Prostatic Dis.*, **17**, 310–314.
38. Bondarev,A.D., Attwood,M.M., Jonsson,J., Chubarev,V.N., Tarasov,V.V. and Schiöth,H.B. (2021) Recent developments of HDAC inhibitors: emerging indications and novel molecules. *Br. J. Clin. Pharmacol.*, **87**, 4577–4597.
39. Spagnuolo,C., Russo,G.L., Orhan,I.E., Habtemariam,S., Daglia,M., Sureda,A., Nabavi,S.F., Devi,K.P., Loizzo,M.R., Tundis,R., *et al.* (2015) Genistein and cancer: current status, challenges, and future directions. *Adv. Nutr.*, **6**, 408–419.
40. Fang,J., Seki,T. and Maeda,H. (2009) Therapeutic strategies by modulating oxygen stress in cancer and inflammation. *Adv. Drug Deliv. Rev.*, **61**, 290–302.

Received: March 25, 2024. **Revised:** August 26, 2024. **Editorial Decision:** October 28, 2024. **Accepted:** October 30, 2024