Preprints.org

Review

An Updated Overview of Existing Cancer Databases and Identified Needs

Brittany Austin , Ali Firooz , Homayoun Valafar * , Anna V. Blenda *

Review

An Updated Overview of Existing Cancer Databases and Identified Needs

Brittany Austin [1], Ali Firooz [2], Homayoun Valafar [2],* and Anna V. Blenda [1,3,*]

[1] USC School of Medicine Greenville; bkaustin@email.sc.edu
[2] USC College of Engineering and Computing; ali.firooz@sc.edu; homayoun@cse.sc.edu
[3] Prisma Health Cancer Institute
* Correspondence: ablenda@greenvillemed.sc.edu

Simple Summary: This review examines the current state of cancer databases and identifies key needs in the field. The analysis of 71 databases reveals a lack of dedicated lipidomic and glycomic databases for cancer research, as well as limited proteomic databases. By comparing non-cancer databases, advancements in genomics, proteomics, lipidomics, and glycomics over the past eight years are highlighted. The evaluation of user-friendliness using the FAIRness principle emphasizes the importance of accessibility and usability. Overall, the review emphasizes the growth of cancer databases while identifying areas for improvement, offering valuable insights for researchers, clinicians, and database developers. Addressing these needs will advance cancer research and benefit the wider cancer community.

Abstract: Our search of existing cancer databases aimed to assess the current landscape and identify key needs. We analyzed 71 databases, focusing on genomics, proteomics, lipidomics, and glycomics. We found a lack of cancer-related lipidomic and glycomic databases, indicating a need for further development in these areas. Proteomic databases dedicated to cancer research were also limited. To assess overall progress, we included human non-cancer databases in proteomics, lipidomics, and glycomics for comparison. This provided insights into advancements in these fields over the past eight years. We also analyzed other types of cancer databases, such as clinical trial databases and web servers. Evaluating user-friendliness, we used the FAIRness principle to assess findability, accessibility, interoperability, and reusability. This ensured databases were easily accessible and usable. Our search summary highlights significant growth in cancer databases while identifying gaps and needs. These insights are valuable for researchers, clinicians, and database developers, guiding efforts to enhance accessibility, integration, and usability. Addressing these needs will support advancements in cancer research and benefit the wider cancer community.

Keywords: cancer; database; genomic; proteomic; lipidomic; glycomic; clinical trials

1. Introduction

Cancer has been known for a long time, with credible evidence observed in fossilized dinosaurs and human bones from prehistoric times. The earliest record of cancer, written between 1500 and 1600 BC, was discovered in the 19th century [3]. Great physicians and scholars such as Hippocrates, Celsus, and Galen have contributed to a better understanding of cancer, its origin, and nature [1]. The "modern era" of cancer research began in the 19th century and led to the development of the current understanding by several investigators, notably Rudolf Virchow, who stated that cancer is "a disease of cells" [2]. This marked the onset of the war on cancer [3], with physicians and researchers collecting massive amounts of information about the mechanisms of cancer and its influence on genes, proteins, and other biomolecules.

To aggregate this massive amount of information into a central location, databases shared across the international community of researchers are a must. The availability of these databases plays a crucial role in aiding the discovery of the molecular basis of such a complex disease as cancer. The

first modern cancer databases emerged in the early 1900s as individual physician's or institutional projects in the United States or Europe [4]. It was not until 1959 that the American College of Surgeons (ACoS) formally adopted a policy allowing hospital-based cancer registries (i.e., databases) [4], with the primary importance of those databases for "monitoring cancer incidence, mortality, and survival" [5]. Nowadays, the functionality of cancer databases has significantly expanded through the analysis of complex datasets, including genomic, proteomic, glycomic, and clinical trials, to name a few. This review gives an update on the progress of cancer databases development in the last eight years (2015-2023). Periodic review of the existing cancer databases is needed to identify gaps and needs in our existing data collections and analysis tools. This report is one such example, with a focus on surveying the existing databases that aggregate nucleic acids (various forms of RNA and DNA), proteins, carbohydrates, and lipids in the context of cancer.

2. Materials and Methods

In this literature review focused on cancer databases in genomics, proteomics, lipidomics, and glycomics, our goal is to analyze their development over the past eight years and identify the existing needs within the cancer research community.

To select the databases for inclusion in the manuscript, we applied two criteria. Firstly, we considered databases published after 2015, as a comprehensive review of the human cancer databases was already available prior to that year [108]. However, we did include a number of papers written before 2015, to illustrate the growth and evolution of certain databases over time. Secondly, we ensured that the selected databases were cancer related. Following these criteria, we compiled a list of 95 databases covering multiple areas of cancer research. From this list, we decided to focus on genomics, proteomics, lipidomics, and glycomics as the fields of interest.

During our analysis, we observed the absence of cancer-related lipidomic and glycomic databases, and only a few cancer-related proteomic databases. Consequently, we decided to incorporate several human non-cancer databases that contain proteomic, lipidomic, and glycomic data. This allowed us to compare the overall progress of knowledge in these fields over the last eight years (2015-2023).

Furthermore, we examined other types of cancer databases, including databases of cancer clinical trials, web servers, and various other cancer-related databases that did not fit into the aforementioned categories. In total, our final selection comprised 71 databases, consisting of 26 genomic, 10 proteomic, 2 lipidomic, 13 glycomic, 7 dedicated to clinical trials, 6 web servers, and 9 other databases. Out of these, 46 databases were cancer-related, while 25 were human non-cancer-related. For our analysis, we utilized 108 sources, primarily published after 2015, including 101 original articles and 7 website sources. Additionally, 40 sources were published before 2015, while 61 sources were published after that year.

Finally, we applied the FAIRness principle to evaluate the user-friendliness of the databases. The FAIR principle emphasizes that databases should be findable, accessible, interoperable, and reusable. To assess these criteria, we conducted our own research on each database. If a database was easily discoverable through web browsers such as Google or Safari, it was considered findable. If the database allowed for login or free access, it was considered accessible. Interoperability was determined by the presence of the database's own statistical analysis function. Lastly, a database was considered reusable if it provided users with the ability to download data. ChatGPT technology was used at the last stage of the revision process of the manuscript.

3. Results

3.1. Genomic Databases

Genetic mutations are pivotal in cancer development, and the National Institute of Health (NIH) established the Cancer Genome Atlas (TCGA) to identify significant cancer-causing genomic changes. TCGA has amassed over 11,000 cases spanning 33 tumor types, providing a vast dataset of molecular alterations [6]. Other databases have leveraged TCGA data, such as the OncomiR Cancer

Database (OMCD), which utilizes TCGA's 9,500 cancer tissue samples for comparative genomic analyses of miRNA sequencing data [7]. Similarly, Cistrome Cancer serves as a web-based server utilizing TCGA to facilitate data retrieval for integrative gene regulation modeling [8]. Notably, there is a trend of creating smaller user-friendly databases derived from larger ones, exemplified by the cBio Cancer Genomic Portal. Developed to integrate extensive genomic projects, cBio enhances accessibility of raw data to the cancer research community [9].

The International Cancer Genome Consortium (ICGC) is another database aiming to construct a comprehensive catalog of mutational abnormalities observed in major tumor types [10]. ICGC incorporates data from 84 global cancer projects, encompassing approximately 77 million somatic mutations and molecular data from over 20,000 participants [10]. The Human Genome Browser at UCSC acts as a portal for displaying various genomic features, including gene predictions, alignments, polymorphisms, and more [11,12]. The Gene Expression Omnibus Database (GEO), established in 2000, focuses on gene expression and functional genomic datasets, extending beyond genome analysis to genome methylation, chromatin structure, and more [13]. Ensembl, created by Flicek et al. in 2014, provides tools for genomic analysis and has expanded each year. In Ensemble 2018, fields like gene annotation, comparative genomics, genetics, and epigenomics were added by Zerbino et al. [14,15]. Recently, Martin et al. expanded Ensemble's genome analysis beyond humans to investigate pangenomes across diverse species in Ensemble 2023 [16].

The Roche Cancer Genome Database 2.0 (RCGDB) serves as a comprehensive platform that combines different human mutation databases into a single location. This database offers interactive search capabilities for genes, samples, cell lines, diseases, and pathways, providing users with a centralized resource for accessing and analyzing cancer-related information. RCGDB also allows for customized searches based on specific filter criteria, enabling researchers to address regularly occurring queries efficiently [17]. The National Cancer Institute Genomic Data Commons (GDC) is another prominent cancer database that focuses on storing, analyzing, and sharing genomic and clinical data from cancer patients. The GDC aims to democratize access to cancer genomic data and promote data sharing among researchers. By facilitating the application of precision medicine approaches, the GDC contributes to advancing the diagnosis and treatment of cancer [18,19]. OpenGDC, derived from the GDC, expands upon the existing platform by incorporating the Genomic Data Model. It introduces additional genomic data in Browser Extensible Data (BED) format and provides related metadata in a table-limited key-value format. OpenGDC enhances the efficiency of accessing genomic and clinical data while expanding the amount of information available for analysis [20].

A notable trend observed in cancer databases is the integration of diverse areas of cancer research into a single platform, allowing for the incorporation of multiple functionalities within a unified database. The Gene Expression Omnibus Database (GEO) serves as an example of such integration, offering not only gene expression data but also functional genomic datasets related to genome methylation, chromatin structure, and genome analysis. By encompassing various aspects of cancer research, GEO facilitates comprehensive investigations and analysis within a single database [13].

Futreal et al. emphasize the importance of mutations occurring in more than 1% of genes in the context of human cancers [21]. To facilitate easy access to information about these genes for researchers and physicians, several databases and web servers focus on cataloging them. Examples of such databases include the Network of Cancer Genes [22], CancerGenes [23], and Cancer Hallmark Genes (CHG) [24]. These databases specifically examine genes that are significantly impacted or mutated in cancer.

The Catalogue of Somatic Mutations in Cancer (COSMIC) database is another valuable resource that stores somatic mutation data and related information about human cancer [25]. Since 2004, COSMIC has integrated coding mutations into its database, covering various genetic mechanisms through which somatic mutations contribute to cancer development. These mechanisms include non-coding mutations, gene fusions, copy-number variants, and drug resistance mutations [26]. Additionally, the COSMIC website provides users with the ability to visualize the 3D structure of proteins [26].

4

Mutagene is a database that delves into the mutational profiles of 37 distinct cancer types. It investigates the underlying components and signatures across over 9,000 genomes and exomes, enabling comparisons of mutagenic processes between different types of cancers [27]. The Progenetix project, initiated in 2001, focuses on individual cancer copy number abnormalities (CNAs) profiles and associated metadata. Over the years, the project has expanded its collection of copy number variations (CNVs) and increased the number of samples, resulting in an improved database with enhanced data quality [28,29]. The MutEx database is dedicated to gathering information on the connections between somatic mutations, gene expression, and patient survival rates [30].

Oncomine is a cancer microarray database that conducts genome-wide expression analyses to identify tumor-related genes, novel biomarkers, and therapeutic targets [31]. Oncomine 3.0, developed in 2007, serves the biomedical research community by collecting, standardizing, analyzing, and delivering cancer transcriptome data [32]. Rhodes et al. utilized the Oncomine 3.0 database to identify genes, pathways, cancer types, and subtypes [32]. Currently, Oncomine has focused its efforts on assay analysis to assist oncologists in making clinical decisions. Their latest functional version is Oncomine Comprehensive Assay v3 (OCAv3), which covers 151 cancer-associated genes, allowing the detection of single nucleotide variants (SNVs), multiple-nucleotide variants (MSVs), and small insertions/deletions (indels) [33]. Since 2017, OCAv3 has been used in clinical settings to support oncologists in determining therapeutic courses. Additionally, Oncomine has developed Oncomine Comprehensive Assay Plus (OCA-Plus), which covers 501 genes, with 144 genes overlapping with OCAv3. OCA-Plus includes assays for microsatellite instability (MSI) and tumor mutational burden (TMB), all in one workflow. Currently, the update of OCA-Plus is under development before its release into clinical settings [33].

3.1.1. CancerResource

The CancerResource database is a comprehensive cancer-related data repository that integrates information from multiple databases to provide a fuller and more interactive resource. One key aspect of CancerResource is its focus on understanding how medications or drug-related substances interact with specific genes or proteins [34]. To achieve its comprehensive approach, CancerResource utilizes several databases, including the Comparative Toxicogenomic Database (CTD), Therapeutic Target Database (TTD), Pharmacogenomics Knowledge Base (PharmGKB), and DrugBank. The Comparative Toxicogenomic Database (CTD) connects toxicological data related to chemicals, genes, phenotypes, diseases, and exposures to enhance our understanding of human health [35]. The Therapeutic Target Database (TTD) provides information on known therapeutic proteins and nucleic acid targets. It includes pathway information and details about drugs/ligands directed at each target. The database offers sequences, 3D structures, functions, nomenclature, drug/ligand binding properties, drug usage, and effects associated with each target. Over time, TTD has expanded its repository to include target-regulating microRNAs, transcription factors, target-interacting proteins, as well as patented agents and their corresponding targets [36,37]. The Pharmacogenomics Knowledge Base (PharmGKB) presents genotypes, molecular data, and clinical information in a pathway-oriented representation. It also provides Very Important Pharmacogenes (VIP) summaries and links to additional external sources for further exploration. As of April 2021, PharmGKB contained annotated data for 715 drugs, 1,761 genes, 227 diseases, and 165 clinical guidelines and drug labels [38,39]. DrugBank is a database that offers detailed molecular information about medications, including mechanisms, interactions, and targets. The most recent edition is DrugBank 5.0 [40].

In the last eight years, the CancerResource database has expanded, encompassing approximately 91,000 drug-target relations, over 2,000 cancer cell lines, and drug sensitivity data for about 50,000 drugs. CancerResource also allows users to upload external expression and mutation data, enabling comparison with the database's cell lines [41]. It is worth noting that as individual databases grow, interconnected databases like CancerResource benefit from the acquisition of new and valuable information.

3.1.2. Cancer Specific Databases

Lung Explore (LCE) is a database specifically dedicated to lung cancer. It enables researchers and clinicians to explore lung cancer data and perform various analyses [42]. PROMISE (Prostate Cancer Precision Medicine Multi-Institutional Collaborative Effort) is a consortium that aims to establish a collection of de-identified clinical and genomic patient data linked to patient outcomes. PROMISE involves different committees focusing on genomic data, statistical analyses, patient advocacy, and other aspects to advance precision medicine in prostate cancer research [43].

HCCDB is a notable database that focuses on hepatocellular carcinoma (HCC), a type of liver cancer. It serves as an online resource providing a consolidated platform for researching gene expression in relation to HCC. HCCDB allows for different types of analyses, including tissue-specific and tumor-specific expression analysis, as well as co-expression analysis [44].

OncoReveal database specifically focuses on non-small cell lung cancer (NSCLC) and colorectal cancer (CRC) [45]. It provides a platform for researchers and clinicians to access relevant data and insights related to these specific cancer types. For a summary of all the GENOMIC databases and web servers reviewed, as well as a visual representation of the information, please refer to Figure 1 and Table 1.
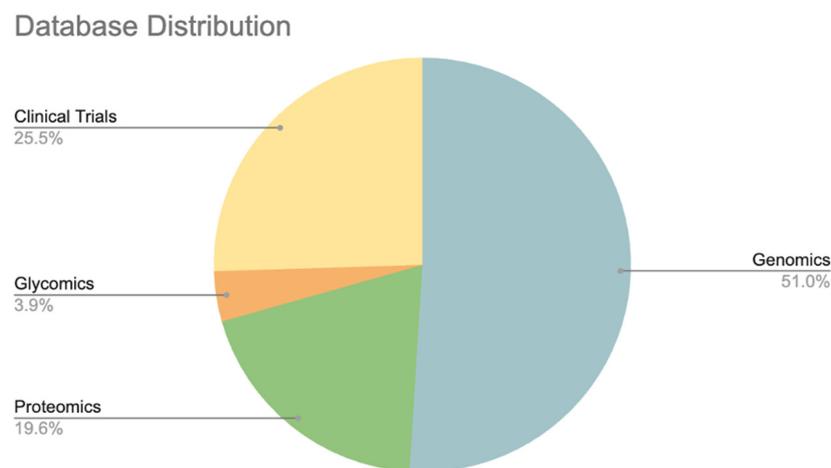
Figure 1. Distribution of Molecular Databases. The pie chart illustrates the distribution of molecular databases among the different types. A total of 51 databases were included in the analysis, comprising 26 Genomic databases, 10 Proteomic databases, 2 Lipidomic databases, and 13 Glycomic databases.

Table 1. Genomic Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
The Cancer Genome Atlas Cases= 11,315	Genome sequencing across 33 tumor types	Yes	Yes	Yes	F, A, I, R	https://www.cancer.gov/ccg/research/genome-sequencing/tcga
OncomiR Cancer Database OMCD Cases= 9500	Comparative genomic analysis of miRNA data sequencing	Yes	N/A	Yes	F, A, I	http://www.oncomir.org/cgi-bin/dbSearch.cgi
cBio Cancer Genomic Portal	Genomic analysis of cancer-related genes	Yes	Yes	Yes	F, A, I, R	https://www.cbioportal.org/
International Cancer Genome Consortium (ICGC) Donors~ 24,500	Catalog of mutational abnormalities in the major tumor types	Yes	Yes	Yes	F, A, I, R	https://dcc.icgc.org/
Human Genome Browser at USCS	Genomic data	Yes	Yes		F, A, R	https://genome.ucsc.edu/index.html
Gene Expression Omnibus Database (GEO)	Gene expression data	Yes	Yes		F, A, R	https://www.ncbi.nlm.nih.gov/geo/
Ensembl	Genomic analysis	Yes	Yes		F, A, R	https://www.ensembl.org/index.html
Roche Cancer Genome Database (RCGDB)						
National Cancer Institute Genomic Commons (GDC) Cases= 22,000	Storage, analysis, and sharing of clinical data of patients	Yes	Yes	Yes	F, A, I, R	https://portal.gdc.cancer.gov/
Network of Cancer Genes	Cancer genes, healthy drivers and their properties	Yes	Yes	Yes	F, A, I, R	http://ncg.kcl.ac.uk/index.php
CancerGenes	Could not find					
Catalogue of Somatic Mutation in Cancer (COSMIC)	Genetic mechanisms that promote cancer	Yes	Yes	Yes	F, A, I, R	https://cancer.sanger.ac.uk/cosmic
Mutagene	Mutational profiles in 37 cancer types	Yes	Yes	Yes	F, A, I, R	https://www.ncbi.nlm.nih.gov/research/mutagene/

Progenetix Samples= 142,063	Cancer Copy Number Abnormalities (CNA)	Yes	Yes	Yes	F, A, I, R	https://progenetix.org/
MutEx	Records the relationships between gene expression, somatic mutation, and survival data	Yes	Yes	Yes	F, A, I, R	http://www.innovebioinfo.com/Databases/Mutationdb_About.php
Oncomine	Precision oncology	Yes	Yes	Yes	F, A, I, R	https://www.oncomine.com/
CancerResource	Server taken down					
Comparative Toxicogenomic Database (CTD)	Toxicological information	Yes	Yes	Yes	F, A, I, R	http://ctdbase.org/
Therapeutic Target Database (TTD)	Pathway information and the drug/ligands directed at each target	Yes	Yes	Yes	F, A, I, R	https://db.idrblab.net/ttd/
Pharmacogenomics Knowledge Base (PharmGKB)	Genotype, molecular, and clinical knowledge integrated into pathway representation	Yes	Yes		F, A, I, R	https://www.pharmgkb.org/
DrugBank	Molecular information about drugs, mechanisms, and interactions	Yes	Yes		F, A, I, R	https://go.drugbank.com/
Lung Cancer Explore (LCE) Entries= 356	Molecular information about drugs including interactions and targets	Yes	Yes		F, A, R	https://lce.biohpc.swmed.edu/lungcancer/imageset_tcga.php
Prostate Cancer Precision Medicine Multi-Institutional Collaborative Effort PROMISE	DNA kit, analyzes genes and patient outcomes	Yes		Yes	F, A, I	https://www.prostatecancerpromise.org/research/
HCCDb	Contains information on hepatocellular carcinoma	Yes	Yes		F, A, R	http://lifeome.net/database/hccdb/home.html
HCCDb	Contains information on hepatocellular carcinoma	Yes	Yes		F, A, R	http://lifeome.net/database/hccdb/home.html

*Databases with names in bold are non-cancer related.

3.2. Proteomic Databases

The Clinical Proteomic Tumor Analysis Consortium (CPTAC) is a database created by the National Cancer Institute (NCI) that analyzes cancer biospecimens using mass spectrometry. It identifies and characterizes protein alterations within tumor samples, providing this proteomic data to the public in an accessible manner. CPTAC collaborates with the Cancer Genome Atlas (TCGA) to provide proteomic input for breast, colorectal, and ovarian tissue samples within the TCGA framework [46,47]. Lindgren's paper in 2021 discusses the data application programming interface (API) created by CPTAC, which distributes processed data sets in a consistent format, facilitating advanced analysis [48].

The String database integrates known and predicted associations between proteins, including physical interactions and functional associations. It utilizes text mining, pathway analysis, and interaction databases to consolidate knowledge on protein interactions [49].

The UALCAN web portal, established in 2017, allows the cancer community to analyze and access cancer transcriptome, proteomics, and patient survival data. It has been expanded to include microRNAs, long non-coding RNAs (lncRNAs), DNA methylation data, and proteomics from CPTAC [50].

CanProVar focuses on human cancer proteome variations, providing a platform for the storage and retrieval of single amino acid alterations observed in cancer. Researchers can efficiently query and explore these alterations using CanProVar, which offers easy accessibility and search capabilities based on gene or protein IDs, cancer types, chromosome locations, and pathways. CanProVar 2.0 is the latest version, featuring a tenfold increase in the number of variations and improved search functionality [51].

The following resources mentioned below are not specifically cancer-related, but they contribute to the understanding of proteomics and its role in cancer research. The RCSB Protein Data Bank provides access to 3D structures of biological macromolecules, aiding in the comprehension of protein and macromolecule structures [52]. The Universal Protein Resource (UniProt) is an open-source repository of protein sequences and functional annotations, offering visualizations of protein subcellular localization, structure, and interactions [53,54]. Proteome Discoverer is a data software used to convert mass spectrometry files to protein identifications [55]. SWISS-PROT and TrEMBL are protein sequence databases that provide information on protein functions, domains, structures, and post-translational modifications [56]. jPOST is a proteomic database that allows users to observe the frequency of post-translational modification detection, examine the co-occurrence of phosphorylation sites, and explore peptide sharing among proteoforms [57]. MatrisomeDB is a selected proteomic database containing data from various extracellular matrix (ECM) studies, offering a searchable repository of useful information related to normal tissues, cancers, and disorders [58]. Table 2 provides a summary of the mentioned proteomic databases.

Table 2. Proteomic Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	Analyzes cancer biospecimens using mass spectrometry	Yes	Yes	Yes	F, A, I, R	https://proteomics.cancer.gov/programs/cptac
String Database	Protein interactions	Yes	Yes	Yes	F, A, I, R	https://string-db.org/
Ualcan	Analyzes and delivers cancer transcriptome, proteomics, and patient survival	Yes	N/A	Yes	F, A, I	https://ualcan.path.uab.edu/
CanProVar	Proteomic variations	Yes	Yes		F, A, R	http://119.3.70.71/CanProVar/index.html
RCSB Protein Data Bank	Works with UniProt and looks at structures of proteins	Yes	Yes	Yes	F, A, I, R	https://www.rcsb.org/
Universal Protein Resource (UniProt)	Contains protein structure and interaction	Yes	Yes	Yes	F, A, I, R	https://www.uniprot.org/
Proteome Discover	Not free to access					
Swiss-Prot and TrEMBL	A part of the UniProt database	Yes	Yes	Yes	F, A, I, R	https://www.uniprot.org/uniprotkb?query=%2A
jPOST	Post-translational modifications on proteins	Yes	Yes	Yes	F, A, I, R	https://globe.jpostdb.org/
MatrisomeDB	Proteomic data from studies on ECM	Yes	Yes		F, A, R	https://matrisomedb.org/

* Databases with names in bold are non-cancer related.

3.3. Lipidomics

Lipidomics plays an increasingly important role in cancer research due to the involvement of lipids in cancer growth, including their role in membrane structure, energy storage, and signal transduction. Some cancer cells, such as breast and ovarian cancer cells, rely on fatty acid oxidation for energy, while lipid accumulation has been observed in certain cancer cells [59]. Understanding the specific lipids affected in different types of cancer can aid in the development of improved treatments and diagnostic approaches.

Although lipidomics in cancer research is still under development, studies have explored the role of lipids in various cancers. For example, a study on lipidomics in colorectal cancer suggested that lipids may play a role in cancer development. However, further research involving larger populations and different cancer stages is needed. Additionally, investigating other factors contributing to increased lipid production in cancer cells is recommended [60].

While there is currently no cancer-specific lipidomics database, there are non-cancer lipidomic databases that provide valuable resources (Table 3). One such database is DBLiPro, which aims to establish a comprehensive knowledge base of human lipid metabolism and offers lipidome-centric analysis tools [61]. Lipid Maps is another notable database, consisting of two components: the Lipid Maps Proteome database (LMPD), which focuses on proteins [62], and the Lipid Maps Structure database (LMSD), which provides information on lipid structures and annotations of biologically relevant lipids [63]. In 2020, Lipid Maps updated its classification system and shorthand notation for lipid structures, including categories such as fatty acyls and glycerolipids [64].

Table 3. Lipidomic Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
DBLiPro		Yes	Yes	Yes	F, A, I, R	http://lipid.cloudn a.cn/home
Lipid Maps		Yes	Yes	Yes	F, A, I, R	https://www.lipid maps.org/

* Databases with names in bold are non-cancer related.

3.4. Glyco Databases

Galectin and glycomic research have gained importance in cancer studies due to their involvement in crucial processes like angiogenesis, metastasis, cell division, and immune evasion. Specific galectins and glycans play significant roles in these processes, modifying immune cells through interactions with glycosylated proteins and lipids. Understanding the effects of galectins and glycans and their alterations in cancer can lead to improved diagnostics, treatments, and drugs. Changes in galectin expression may be influenced by protein trafficking and alterations in the glycocalyx composition of cancer cells [65–68].

While most glycomic databases are not cancer-specific, they provide valuable insights into glycan structure, function, and the field of glycoproteomics. Glycoproteomics focuses on identifying, locating, characterizing, and studying the abundance and role of glycosylated proteins in biological processes, including cancer. Mass spectrometry is commonly used for studying glycan alterations in cancer [69–73].

Given the limited number of cancer-related glycomic databases, incorporating glycomic information into cancer-related databases is crucial. Key glycomic databases include GlycoSuiteDB, UniCarb-DB, EUROCarbDB, UniPep, GlycoGene database (GGDB), Glycome-DB, and Glyco-base. These databases offer a wealth of glycan and glycoproteomics data, enabling the examination of glycan structures, fragment data, biological context, and more [74-80]. Recent advancements in the field include GlycoStore, GlycoRDF, GRITs database, GlyTouCan, Lectin Frontier Database (LfDB), and Carbohydrate Structure Database (CSDB), aiming to improve data quality, coverage, and standardization of carbohydrate notations [81-86], (Table 4, Figure 2).

Table 4. Glyco Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
GlycoSuite	Was not able to find					
UniCarb-db	Carbohydrates characterized by LC-MS	Yes	Yes	Yes	F, A, I, R	https://unicarb-db.expasy.org/
EuroCarbDB	Was not able to find					
UniPep	N-linked glycosites for proteomic analyses	Yes	Yes		F, A, R	https://unipep.systemsbiology.net/
GlycoGene (GGDB)	Contains all the information on glycogenes	Yes	Yes		F, A, R	https://www.glycogene.com/
Glycome-DB	A part of GlyTouCan database					http://www.glycome-db.org/
Glycobase	Was not able to find					
GlycoStore	Was not able to find					
GlycoRDF	Holds glycan publications and experimental data	Yes	Yes		F, A, R	https://github.com/glycoinfo/GlycoRDF/wiki
GRITs Toolbox	Allows for archiving of research papers	Yes	Yes	Yes	F, A, I, R	http://www.grits-toolbox.org/
GlyTouCan	Databases for publications and journals within glycan research	Yes	Yes		F, A, R	https://glytoucan.org/
The Lectin Frontier Database (LfDB)	Lectin-standard oligosaccharide interactions	Yes			F, A	https://acgg.asia/lfdb2/
Carbohydrate Structure Database (CSDB)	Structural and biographical components of glycans	Yes		Yes, this is done by an external source	F, A	http://csdb.glycoscience.ru/database/index.html?help=credits

*Databases with names in bold are non-cancer related.
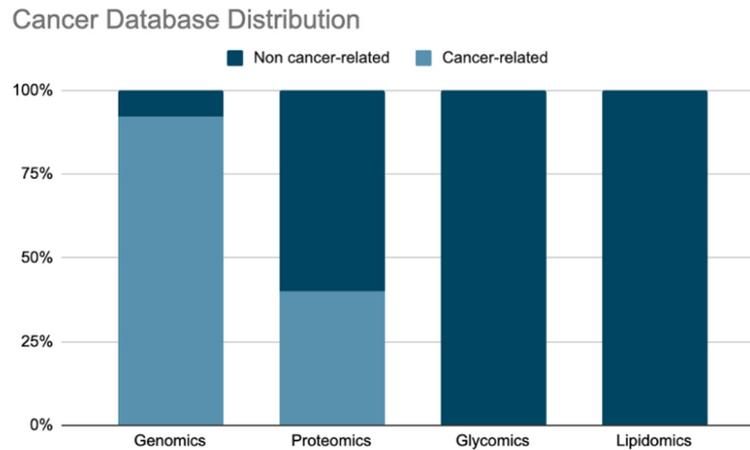
Figure 2. Distribution of Cancer-Related and Non-Cancer-Related Databases among Molecular Databases.

3.5. Clinical Trial Databases

Clinical trials play a crucial role in cancer research, as they help evaluate the safety and effectiveness of diagnostics, treatments, and medication development. Integrating clinical trial databases is essential for understanding the impact of trials and patient demographics on the development of improved and personalized treatments. Here are several clinical trial databases relevant to cancer research: 1) Clinical Genomic Database (CGD): CGD provides a comprehensive collection of genetic conditions where genetic information can influence appropriate supportive care, medical decision-making, prognostic assessments, reproductive choices, and help avoid unnecessary diagnostic testing [87]. 2) Foundation Medicine Adult Cancer-Clinical Dataset: This dataset serves as a valuable resource for researching uncommon mutations and disorders, verifying their clinical importance, and discovering novel treatment options [88]. 3)Curated Cancer Clinical Outcomes Database (C3OD): C3OD integrates electronic medical records, tumor registry, biospecimen, and data registry to facilitate easier access to patient data in a unified location. Its goal is to accelerate eligibility screening for research purposes [89]. 4) Danish Head and Neck Cancer Database: Started in the early 1960s, this database focuses on a national strategy for multidisciplinary treatment of head and neck cancer in Denmark. It is utilized to describe the effects of reduced waiting time, changing epidemiology, and the influence of comorbidity and socioeconomic factors [90]. 5) National Cancer Database (NCDB): Over the past three decades, NCDB has evolved significantly, aggregating and categorizing approximately 40 million patient records from over 1500 hospitals. Its aim is to enhance the quality of cancer patient care [91]. 6) Surveillance, Epidemiology, and End Results (SEER) database: SEER focuses on investigating the history of colorectal cancer and patient care, providing valuable insights to the field [92]. 7) ClinVar: ClinVar is a public database designed for clinical laboratories, researchers, and expert panels. Launched in 2013, it contains over 600,000 submitted records from 1,000 submitters, representing 430,000 unique variants. ClinVar enables data comparison among researchers [93].

Table 5 includes more detailed information about each database, its main features and scope.

Table 5. Clinical Trial Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
Clinical Genomic Database (GCD)	Genetic information that pertains to patient care	Yes	Yes		F, A, R	https://research.nhgri.nih.gov/CGD/
Foundation Medicine Adult-Cancer-Clinical Dataset	Clinical relevance among rare alterations and diseases	Yes	Yes	Yes	F, A, I, R	https://gdc.cancer.gov/about-gdc/contributed-genomic-data-cancer-research/foundation-medicine/foundation-medicine
A Curated Cancer Clinical Outcome Database (C3OD)	Cannot access					
Danish Head and Neck Cancer Database	Contains patient data to be used in improved wait times	Yes	Yes		F, A, R	https://www.dahanca.dk/IndexPage
National Cancer Database (NCDB)	Have to login for access				F	https://www.facs.org/quality-programs/cancer-programs/national-cancer-database/
Surveillance Epidemiology and End Results (SEER)	Focus on colorectal cancer and improvements of patient care	Yes			F, A	https://seer.cancer.gov/
ClinVar	Allows for the comparison of data among researchers	Yes	Yes		F, A, R	https://www.ncbi.nlm.nih.gov/clinvar/

* Databases with names in bold are non-cancer related.

3.6. Other Cancer Databases

Several other databases are also important for cancer research. The Database of Epigenetics Modifier (dbEM) contains potential targets for cancer treatment and information on mutations, copy number variations, and gene expression in tumor samples [94]. The Cancer Research Database (CRDB) explores the correlation between cancer and the COVID-19 pandemic, scoring other databases based on cancer types, sample size, omics results, and user interface [95]. The Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis discusses databases that examine prognostic biomarkers and survival rates, including PROGgene V2 [96,97]. The Cancer Drug Resistance (CancerDR) database provides information on anti-cancer drugs and their profiling across cancer cell lines [98]. DriverDB identifies driver genes/mutations using algorithms [99], while LncRNA2Target 2.0 and Lnc2Cancer focus on long non-coding RNAs associated with cancer [100,101]. The Genotype-Tissue Expression (GTEx) database investigates the relationship between genetic variation and gene expression in humans [102]. These evolving databases additionally contribute to improved diagnosis, prognosis, and therapeutic interventions in cancer research (Table 6).

3.7. Web-based Servers

Web servers are instrumental in cancer research, offering various functionalities and benefits. GSCALite, for example, performs comprehensive analysis of cancer-related genes, including differential expression, survival analysis, genomic variation assessment, cancer pathway activity, miRNA regulation, drug sensitivity, and normal tissue expression [103]. OMIM serves as an online catalog, providing extensive information on genetic phenotypes, DNA/protein sequences, references, and mutational databases [104]. GEPIA is a web-based tool that enables interactive analysis of differential gene expression, correlation, survival, gene similarity, and dimensional reduction [105]. PepQuery facilitates proteomic validation of genomic alterations through simulations and experimental data [106]. These web servers play a critical role in empowering researchers and enabling in-depth exploration and analysis of cancer data (Table 7).

Table 6. Other Databases.

Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
Database of Epigenetics Modifiers (dbEM)	Contains genomic information on epigenetic modifiers/ proteins	Yes			F, A	https://webs.iiitd.edu.in/raghava/dbem/index.php
Cancer Research Database (CRDB)	Holds other databases in the fields of genomic, proteomic, mutations, etc.	Yes			F, A	https://www.habdsk.org/crdb
PROGgene	Prognosis	Yes		Yes	F, A, I	http://www.progtools.net/gene/index.php
Cancer Drug Resistance (CancerDR)		Yes	Yes		F, A, R	https://webs.iiitd.edu.in/raghava/cancerdr/index.html
DriverDBv3		Yes	Yes	Yes	F, A, I, R	http://driverdb.tms.cmu.edu.tw/
LncRNA2Target 2.0	Was not able to access					
Lnc2Cancers 3.0	Was not able to access					
Genotype Expression Project (GTEx)	Evaluate the relationships between genetic variations and gene expressions	Yes	Yes		F, A, R	https://www.gtexportal.org/home/

*Databases with names in bold are non-cancer related.

Table 7. Web Servers.

| Databases	Content	Web service	Downloadable	Analytics	Fairness	Website
Gene Set Cancer Analysis (GSCALite)	Analyzes gene and survival rates	Yes			F, A	http://bioinfo.life.hust.edu.cn/web/GSCALite/
Online Mendelian Inheritance in Man (OMIM)	Includes multiple resources on genetic phenotype, DNA, proteins, etc.	Yes	Yes		F, A, R	https://www.omim.org/
Gene Expression Profiling Interactive Analysis (GEPIA)	Gene expression analysis, correlations analysis, and patient survival	Yes	Yes	Yes	F, A, I, R	http://gepia.cancer-pku.cn/
PepQuery	Proteomic validations of genomic alterations	Yes	Yes		F, A, R	http://www.pepquery.org/

*Databases with names in bold are non-cancer related.

4. Discussion

Databases have undergone significant growth and development in the past eight years, manifesting in various ways. Firstly, databases have expanded their information by continually adding more data. For instance, CanProVar 2.0 has experienced a tenfold increase in its content since its inception, enabling the dissemination of more comprehensive information. The sharing of data has emerged as a crucial focus for glycomic researchers, leading to the creation of databases such as GlyTouCan and the Carbohydrate Structure database. These databases aim to address integration challenges and other issues prevalent in glycan databases. CancerResource is another exemplar of databases sharing information, as it derives data from multiple sources.

Furthermore, databases have broadened their research scope by incorporating additional topics beyond their original areas of focus. A notable instance is Ualcan, a proteomic database that integrated microRNA and lncRNA data to explore patient survival outcomes. This expansion reflects the inclination of databases to explore diverse research domains within a single platform.

The second aspect of database growth pertains to database design and usability. Database developers and curators have striven to enhance user-friendliness, often evaluated through the FAIRness principle. This principle encompasses various criteria, including findability, accessibility, interpretability, and reusability, to determine the fairness and usability of scientific research, including databases [107]. A user-friendly database should be discoverable, easily accessible, interpretable, and allow data reuse for any purpose. Many databases examined in this study have endeavored to improve user-friendliness through website redesign, resulting in enhanced search engines and capabilities such as copying/pasting or downloading datasets. Additionally, efforts have been made to enable users to create their datasets within the database.

Overall, databases have experienced growth in terms of data expansion and user-friendly design. These advancements facilitate information sharing, enable broader research exploration, and contribute to the usability and accessibility of scientific research databases.
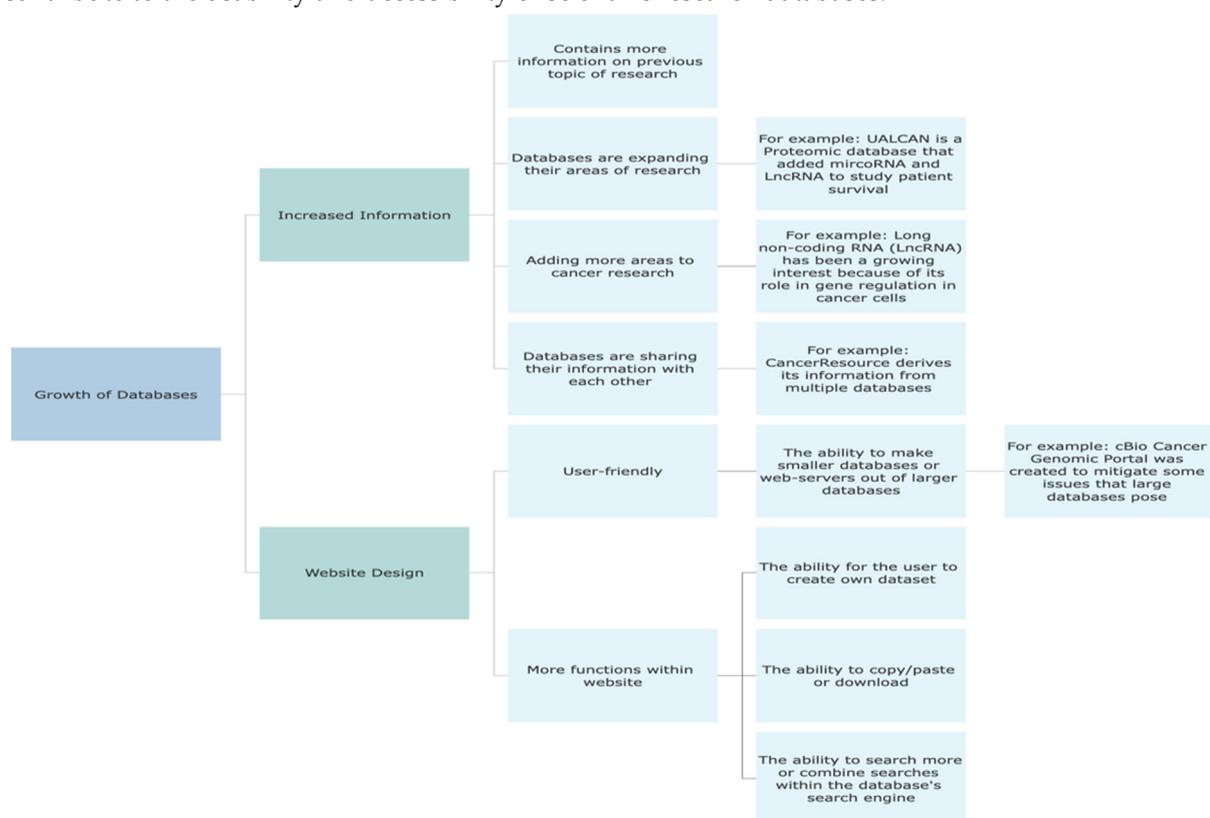
Figure 3. Emerging Trends in Database Development over the Last Eight Years.

5. Conclusions

In conclusion, our search summary of existing cancer databases reveals significant growth and development over the past eight years. We have identified the need for more cancer-related lipidomic and glycomic databases, as well as the scarcity of proteomic databases in the cancer domain. Additionally, we have highlighted the importance of user-friendliness in database design and adherence to the FAIRness principles. This comprehensive analysis provides valuable insights into the current state of cancer databases and the areas that require further attention and improvement.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are available in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   G. B. Faguet, "A brief history of cancer: Age-old milestones underlying our current knowledge database," Int J Cancer, vol. 136, no. 9, pp. 2022–2036, May 2015, doi: 10.1002/IJC.29134.
2.   I. B. Weinstein and K. Case, "The History of Cancer Research: Introducing an AACR Centennial Series," Cancer Res, vol. 68, no. 17, pp. 6861–6862, Sep. 2008, doi: 10.1158/0008-5472.CAN-08-2827.
3.   "SEER Training Modules, Cancer Facts and the War on Cancer," National Cancer Institutes.
4.   "SEER Training Modules, Brief History of Cancer Registration," National Cancer Institute.
5.   G. Ursin, "Cancer registration in the era of modern oncology and GDPR," https://doi.org/10.1080/0284186X.2019.1657586, vol. 58, no. 11, pp. 1547–1548, Nov. 2019, doi: 10.1080/0284186X.2019.1657586.
6.   K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge," Wspolczesna Onkologia, vol. 1A. Termedia Publishing House Ltd., pp. A68–A77, 2015. doi: 10.5114/wo.2014.47136.
7.   A. L. Sarver, A. E. Sarver, C. Yuan, and S. Subramanian, "OMCD: OncomiR Cancer Database," BMC Cancer, vol. 18, no. 1, Dec. 2018, doi: 10.1186/s12885-018-5085-z.
8.   S. Mei et al., "Cistrome cancer: A web resource for integrative gene regulation modeling in cancer," Cancer Res, vol. 77, no. 21, pp. e19–e22, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0327/SUPPLEMENTARY-VIDEO-S1.
9.   E. Cerami et al., "The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data," Cancer Discov, vol. 2, no. 5, pp. 401–404, May 2012, doi: 10.1158/2159-8290.CD-12-0095.
10. J. Zhang et al., "The International Cancer Genome Consortium Data Portal," Nature Biotechnology 2019 37:4, vol. 37, no. 4, pp. 367–369, Mar. 2019, doi: 10.1038/s41587-019-0055-9.
11. W. J. Kent et al., "The Human Genome Browser at UCSC," Genome Res, vol. 12, no. 6, pp. 996–1006, Jun. 2002, doi: 10.1101/GR.229102.
12. "The Human Genome Browser at UCSC." https://genome.cshlp.org/content/12/6/996.short (accessed Feb. 06, 2023).
13. E. Clough and T. Barrett, "The Gene Expression Omnibus database," Methods in Molecular Biology, vol. 1418, pp. 93–110, 2016, doi: 10.1007/978-1-4939-3578-9_5/COVER.

14. P. Flicek et al., "Ensembl 2014," Nucleic Acids Res, vol. 42, no. D1, Jan. 2014, doi: 10.1093/nar/gkt1196.

15. D. R. Zerbino et al., "Ensembl 2018," Nucleic Acids Res, vol. 46, no. D1, pp. D754–D761, Jan. 2018, doi: 10.1093/NAR/GKX1098.

16. F. J. Martin et al., "Ensembl 2023," Nucleic Acids Res, vol. 51, no. D1, pp. D933–D941, Jan. 2023, doi: 10.1093/NAR/GKAC958.

17. J. Küntzer, D. Maisel, H. P. Lenhof, S. Klostermann, and H. Burtscher, "The Roche Cancer Genome Database 2.0," BMC Med Genomics, vol. 4, p. 43, 2011, doi: 10.1186/1755-8794-4-43.

18. M. A. Jensen, V. Ferretti, R. L. Grossman, and L. M. Staudt, "The NCI Genomic Data Commons as an engine for precision medicine," Blood, vol. 130, no. 4, pp. 453–459, Jul. 2017, doi: 10.1182/BLOOD-2017-03-735654.

19. "GDC." https://portal.gdc.cancer.gov/ (accessed Feb. 15, 2023).

20. E. Cappelli et al., "OpenGDC: Unifying, Modeling, Integrating Cancer Genomic Data and Clinical Metadata," Applied Sciences 2020, Vol. 10, Page 6367, vol. 10, no. 18, p. 6367, Sep. 2020, doi: 10.3390/APP10186367.

21. P. A. Futreal et al., "A census of human cancer genes," Nature Reviews Cancer 2004 4:3, vol. 4, no. 3, pp. 177–183, 2004, doi: 10.1038/nrc1299.

22. D. Repana et al., "The Network of Cancer Genes (NCG): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens 06 Biological Sciences 0604 Genetics 11 Medical and Health Sciences 1112 Oncology and Carcinogenesis 06 Biological Sciences 0601 Biochemistry and Cell Biology," Genome Biol, vol. 20, no. 1, pp. 1–12, Jan. 2019, doi: 10.1186/S13059-018-1612-0/FIGURES/4.

23. M. E. Higgins, M. Claremont, J. E. Major, C. Sander, and A. E. Lash, "CancerGenes: a gene selection resource for cancer genome projects," Nucleic Acids Res, vol. 35, no. suppl_1, pp. D721–D726, Jan. 2007, doi: 10.1093/NAR/GKL811.

24. D. Zhang et al., "CHG: A Systematically Integrated Database of Cancer Hallmark Genes," Front Genet, vol. 11, p. 29, Feb. 2020, doi: 10.3389/FGENE.2020.00029/BIBTEX.

25. S. Bamford et al., "The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website," Br J Cancer, vol. 91, no. 2, pp. 355–358, Jul. 2004, doi: 10.1038/sj.bjc.6601894.

26. J. G. Tate et al., "COSMIC: the Catalogue Of Somatic Mutations In Cancer," Nucleic Acids Res, vol. 47, no. D1, pp. D941–D947, Jan. 2019, doi: 10.1093/NAR/GKY1015.

27. A. L. Brown, M. Li, A. Goncearenco, and A. R. Panchenko, "Finding driver mutations in cancer: Elucidating the role of background mutational processes," PLoS Comput Biol, vol. 15, no. 4, 2019, doi: 10.1371/JOURNAL.PCBI.1006981.

28. Q. Huang, P. Carrio-Cordo, B. Gao, R. Paloots, and M. Baudis, "The Progenetix oncogenomic resource in 2021," Database, vol. 2021, no. 0, pp. 1–9, Sep. 2021, doi: 10.1093/DATABASE/BAAB043.

29. "Progenetix." https://progenetix.org/ (accessed Feb. 15, 2023).

30. J. Ping et al., "MutEx: a multifaceted gateway for exploring integrative pan-cancer genomic data," Brief Bioinform, vol. 21, no. 4, pp. 1479–1486, Jul. 2020, doi: 10.1093/BIB/BBZ084.

31. D. R. Rhodes et al., "ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform 1," 2004. [Online]. Available: www.oncomine.org

32. D. R. Rhodes et al., "Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles," Neoplasia, vol. 9, no. 2, pp. 166–180, Feb. 2007, doi: 10.1593/NEO.07112.

33. L. K. Vestergaard, D. N. P. Oliveira, T. S. Poulsen, C. K. Høgdall, and E. V. Høgdall, "OncomineTM comprehensive assay v3 vs. OncomineTM comprehensive assay plus," Cancers (Basel), vol. 13, no. 20, p. 5230, Oct. 2021, doi: 10.3390/CANCERS13205230/S1.

34. J. Ahmed et al., "CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge," Nucleic Acids Res, vol. 39, no. Database issue, p. D960, Jan. 2011, doi: 10.1093/NAR/GKQ910.

35. A. P. Davis et al., "Comparative Toxicogenomics Database (CTD): update 2021," Nucleic Acids Res, vol. 49, no. D1, pp. D1138–D1143, Jan. 2021, doi: 10.1093/NAR/GKAA891.

36. X. Chen, Z. L. Ji, and Y. Z. Chen, "TTD: Therapeutic Target Database," Nucleic Acids Res, vol. 30, no. 1, pp. 412–415, Jan. 2002, doi: 10.1093/NAR/30.1.412.

37. Y. Wang et al., "Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics," Nucleic Acids Res, vol. 48, no. D1, pp. D1031–D1041, Jan. 2020, doi: 10.1093/NAR/GKZ981.

38. C. F. Thorn, T. E. Klein, and R. B. Altman, "PharmGKB: The pharmacogenomics knowledge base," Methods in Molecular Biology, vol. 1015, pp. 311–320, 2013, doi: 10.1007/978-1-62703-435-7_20/COVER.

39. L. Gong, M. Whirl-Carrillo, and T. E. Klein, "PharmGKB, an Integrated Resource of Pharmacogenomic Knowledge," Curr Protoc, vol. 1, no. 8, p. e226, Aug. 2021, doi: 10.1002/CPZ1.226.

40. D. S. Wishart et al., "DrugBank 5.0: a major update to the DrugBank database for 2018," Nucleic Acids Res, vol. 46, no. D1, pp. D1074–D1082, Jan. 2018, doi: 10.1093/NAR/GKX1037.

41. B. O. Gohlke, J. Nickel, R. Otto, M. Dunkel, and R. Preissner, "CancerResource—updated database of cancer-relevant proteins, mutations and interacting drugs," Nucleic Acids Res, vol. 44, no. D1, pp. D932–D937, Jan. 2016, doi: 10.1093/NAR/GKV1283.

42. L. Cai et al., "LCE: an open web portal to explore gene expression and clinical associations in lung cancer," Oncogene 2018 38:14, vol. 38, no. 14, pp. 2551–2564, Dec. 2018, doi: 10.1038/s41388-018-0588-2.

43. V. S. Koshkin et al., "PROMISE: a real-world clinical-genomic database to address knowledge gaps in prostate cancer," Prostate Cancer and Prostatic Diseases 2021 25:3, vol. 25, no. 3, pp. 388–396, Aug. 2021, doi: 10.1038/s41391-021-00433-1.

44. Q. Lian et al., "HCCDB: A Database of Hepatocellular Carcinoma Expression Atlas," Genomics Proteomics Bioinformatics, vol. 16, no. 4, pp. 269–275, Aug. 2018, doi: 10.1016/J.GPB.2018.07.003.

45. "The oncoReveal Dx Lung and".

46. N. J. Edwards et al., "The CPTAC data portal: A resource for cancer proteomics research," J Proteome Res, vol. 14, no. 6, pp. 2707–2713, Jun. 2015, doi: 10.1021/PR501254J/ASSET/IMAGES/MEDIUM/PR-2014-01254J_0003.GIF.

47. "Clinical Proteomic Tumor Analysis Consortium (CPTAC) | NCI Genomic Data Commons." https://gdc.cancer.gov/about-gdc/contributed-genomic-data-cancer-research/clinical-proteomic-tumor-analysis-consortium-cptac (accessed Feb. 06, 2023).

48. C. M. Lindgren et al., "Simplified and Unified Access to Cancer Proteogenomic Data," J Proteome Res, vol. 20, no. 4, pp. 1902–1910, Apr. 2021, doi: 10.1021/ACS.JPROTEOME.0C00919/ASSET/IMAGES/LARGE/PR0C00919_0003.JPEG.

49. D. Szklarczyk et al., "Correction to 'The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets,'" Nucleic Acids Res, vol. 49, no. 18, pp. 10800–10800, Oct. 2021, doi: 10.1093/NAR/GKAB835.

50. D. S. Chandrashekar et al., "UALCAN: An update to the integrated cancer data analysis platform," Neoplasia, vol. 25, pp. 18–27, Mar. 2022, doi: 10.1016/J.NEO.2022.01.001.

51. M. Zhang et al., "CanProVar 2.0: An Updated Database of Human Cancer Proteome Variation," J Proteome Res, vol. 16, no. 2, pp. 421–432, Feb. 2017, doi: 10.1021/ACS.JPROTEOME.6B00505/SUPPL_FILE/PR6B00505_SI_001.PDF.

52. P. W. Rose et al., "The RCSB Protein Data Bank: views of structural biology for basic and applied research and education," Nucleic Acids Res, vol. 43, no. D1, pp. D345–D356, Jan. 2015, doi: 10.1093/NAR/GKU1214.

53. T. U. Consortium, "Activities at the Universal Protein Resource (UniProt)," Nucleic Acids Res, vol. 42, no. 11, pp. 7486–7486, Jun. 2014, doi: 10.1093/NAR/GKU469.

54. A. Bateman, "UniProt: a worldwide hub of protein knowledge," Nucleic Acids Res, vol. 47, no. D1, pp. D506–D515, Jan. 2019, doi: 10.1093/NAR/GKY1049.

55. B. C. Orsburn, "Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics," Proteomes 2021, Vol. 9, Page 15, vol. 9, no. 1, p. 15, Mar. 2021, doi: 10.3390/PROTEOMES9010015.

56. C. O'Donovan, M. J. Martin, A. Gattiker, E. Gasteiger, A. Bairoch, and R. Apweiler, "High-quality protein knowledge resource: SWISS-PROT and TrEMBL," Brief Bioinform, vol. 3, no. 3, pp. 275–284, Sep. 2002, doi: 10.1093/BIB/3.3.275.

57. Y. Moriya et al., "The jPOST environment: an integrated proteomics data repository and database," Nucleic Acids Res, vol. 47, no. D1, pp. D1218–D1224, Jan. 2019, doi: 10.1093/NAR/GKY899.

58. X. Shao, I. N. Taha, K. R. Clauser, Y. (Tom) Gao, and A. Naba, "MatrisomeDB: the ECM-protein knowledge database," Nucleic Acids Res, vol. 48, no. D1, pp. D1136–D1144, Jan. 2020, doi: 10.1093/NAR/GKZ849.

59. F. Yan, H. Zhao, and Y. Zeng, "Lipidomics: a promising cancer biomarker," Clin Transl Med, vol. 7, no. 1, p. e21, Dec. 2018, doi: 10.1186/S40169-018-0199-0.

60. M. Buszewska-forajta et al., "Lipidomics as a diagnostic tool for prostate cancer," Cancers (Basel), vol. 13, no. 9, p. 2000, May 2021, doi: 10.3390/CANCERS13092000/S1.

61. Q. Wu et al., "DBLiPro: A Database for Lipids and Proteins in Human Lipid Metabolism," Phenomics 2023, pp. 1–10, May 2023, doi: 10.1007/S43657-023-00099-W.

62. D. Cotter, A. Maer, C. Guda, B. Saunders, and S. Subramaniam, "LMPD: LIPID MAPS proteome database," Nucleic Acids Res, vol. 34, no. suppl_1, pp. D507–D510, Jan. 2006, doi: 10.1093/NAR/GKJ122.

63. M. Sud et al., "LMSD: LIPID MAPS structure database," Nucleic Acids Res, vol. 35, no. suppl_1, pp. D527–D532, Jan. 2007, doi: 10.1093/NAR/GKL838.

64. G. Liebisch et al., "Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures," J Lipid Res, vol. 61, no. 12, pp. 1539–1555, Dec. 2020, doi: 10.1194/jlr.S120001025.

65. B. B. Blair et al., "Increased circulating levels of galectin proteins in patients with breast, colon, and lung cancer," Cancers (Basel), vol. 13, no. 19, Oct. 2021, doi: 10.3390/cancers13194819.

66. S. S. Pinho and C. A. Reis, "Glycosylation in cancer: mechanisms and clinical implications," Nature Reviews Cancer 2015 15:9, vol. 15, no. 9, pp. 540–555, Aug. 2015, doi: 10.1038/nrc3982.

67. F. T. Liu and S. R. Stowell, "The role of galectins in immunity and infection," Nature Reviews Immunology 2023, pp. 1–16, Jan. 2023, doi: 10.1038/s41577-022-00829-7.

68. A. T. Funkhouser et al., "KIT Mutations Correlate with Higher Galectin Levels and Brain Metastasis in Breast and Non-Small Cell Lung Cancer," Cancers (Basel), vol. 14, no. 11, Jun. 2022, doi: 10.3390/cancers14112781.

69. D. B. Hizal et al., "Glycoproteomic and glycomic databases," Clin Proteomics, vol. 11, no. 1, pp. 1–10, Apr. 2014, doi: 10.1186/1559-0275-11-15/FIGURES/2.

70. Y. Tian and H. Zhang, "Glycoproteomics and clinical applications," Proteomics Clin Appl, vol. 4, no. 2, pp. 124–132, Feb. 2010, doi: 10.1002/PRCA.200900161.

71. E. H. Kim, & D. E. Misek, "Glycoproteomics-based identification of cancer biomarkers," Int J Proteomics, 2011.

72. S. Pan, R. Chen, R. Aebersold, and T. A. Brentnall, "Mass Spectrometry Based Glycoproteomics—From a Proteomics Perspective *," Molecular & Cellular Proteomics, vol. 10, no. 1, p. R110.003251, Jan. 2011, doi: 10.1074/MCP.R110.003251.

73. J. A. Ferreira, M. Relvas-Santos, A. Peixoto, A. M.N. Silva, and L. Lara Santos, "Glycoproteogenomics: Setting the Course for Next-generation Cancer Neoantigen Discovery for Cancer Vaccines," Genomics, Proteomics and Bioinformatics, vol. 19, no. 1. Beijing Genomics Institute, pp. 25–43, Feb. 01, 2021. doi: 10.1016/j.gpb.2021.03.005.

74. C. A. Cooper, M. J. Harrison, M. R. Wilkins, and N. H. Packer, "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources," Nucleic Acids Res, vol. 29, no. 1, pp. 332–335, Jan. 2001, doi: 10.1093/NAR/29.1.332.

75. C. A. Hayes et al., "UniCarb-DB: a database resource for glycomic discovery," Bioinformatics, vol. 27, no. 9, pp. 1343–1344, May 2011, doi: 10.1093/BIOINFORMATICS/BTR137.

76. C. W. Von Der Lieth et al., "EUROCarbDB: An open-access platform for glycoinformatics," Glycobiology, vol. 21, no. 4, pp. 493–502, Apr. 2011, doi: 10.1093/GLYCOB/CWQ188.

77. H. Zhang et al., "UniPep - A database for human N-linked glycosites: A resource for biomarker discovery," Genome Biol, vol. 7, no. 8, pp. 1–12, Aug. 2006, doi: 10.1186/GB-2006-7-8-R73/FIGURES/4.

78. A. Togayachi, K.-Y. Dae, T. Shikanai, and H. Narimatsu, "A Database System for Glycogenes (GGDB)," Experimental Glycoscience, pp. 423–425, Mar. 2008, doi: 10.1007/978-4-431-77922-3_102.

79. R. Ranzinger, M. Frank, C. W. Von der lieth, and S. Herget, "Glycome-DB.org: A portal for querying across the digital world of carbohydrate sequences," Glycobiology, vol. 19, no. 12, pp. 1563–1567, Dec. 2009, doi: 10.1093/GLYCOB/CWP137.

80. M. P. Campbell, L. Royle, C. M. Radcliffe, R. A. Dwek, and P. M. Rudd, "GlycoBase and autoGU: tools for HPLC-based glycan analysis," Bioinformatics, vol. 24, no. 9, pp. 1214–1216, May 2008, doi: 10.1093/BIOINFORMATICS/BTN090.

81. S. Zhao et al., "GlycoStore: a database of retention properties for glycan analysis," Bioinformatics, vol. 34, no. 18, pp. 3231–3232, Sep. 2018, doi: 10.1093/BIOINFORMATICS/BTY319.

82. R. Ranzinger et al., "GlycoRDF: an ontology to standardize glycomics data in RDF," Bioinformatics, vol. 31, no. 6, pp. 919–925, Mar. 2015, doi: 10.1093/BIOINFORMATICS/BTU732.

83. D. B. Weatherly, F. S. Arpinar, M. Porterfield, M. Tiemeyer, W. S. York, and R. Ranzinger, "GRITS Toolbox—a freely available software for processing, annotating and archiving glycomics mass spectrometry data," Glycobiology, vol. 29, no. 6, pp. 452–460, Jun. 2019, doi: 10.1093/GLYCOB/CWZ023.

84. M. Tiemeyer et al., "GlyTouCan: an accessible glycan structure repository," Glycobiology, vol. 27, no. 10, pp. 915–919, Oct. 2017, doi: 10.1093/GLYCOB/CWX066.

85. J. Hirabayashi, H. Tateno, T. Shikanai, K. F. Aoki-Kinoshita, and H. Narimatsu, "The Lectin Frontier Database (LfDB), and Data Generation Based on Frontal Affinity Chromatography," Molecules 2015, Vol. 20, Pages 951-973, vol. 20, no. 1, pp. 951–973, Jan. 2015, doi: 10.3390/MOLECULES20010951.

86. P. V. Toukach and A. I. Shirkovskaya, "Carbohydrate Structure Database and Other Glycan Databases as an Important Element of Glycoinformatics," Russ J Bioorg Chem, vol. 48, no. 3, pp. 457–466, Jun. 2022, doi: 10.1134/S1068162022030190/METRICS.

87. B. D. Solomon, A. D. Nguyen, K. A. Bear, and T. G. Wolfsberg, "Clinical genomic database," Proc Natl Acad Sci U S A, vol. 110, no. 24, pp. 9851–9855, Jun. 2013, doi: 10.1073/PNAS.1302575110/SUPPL_FILE/SD01.XLSX.

88. R. J. Hartmaier et al., "High-throughput genomic profiling of adult solid tumors reveals novel insights into cancer pathogenesis," Cancer Res, vol. 77, no. 9, pp. 2464–2475, May 2017, doi: 10.1158/0008-5472.CAN-16-2479/657735/AM/HIGH-THROUGHPUT-GENOMIC-PROFILING-OF-ADULT-SOLID.

89. D. P. Mudaranthakam et al., "A Curated Cancer Clinical Outcomes Database (C3OD) for accelerating patient recruitment in cancer clinical trials," JAMIA Open, vol. 1, no. 2, pp. 166–171, Oct. 2018, doi: 10.1093/JAMIAOPEN/OOY023.

90. J. Overgaard, A. Jovanovic, C. Godballe, and J. Grau Eriksen, "The Danish Head and Neck Cancer database," Clin Epidemiol, vol. 8, pp. 491–496, Oct. 2016, doi: 10.2147/CLEP.S103591.

91. R. M. McCabe, "National Cancer Database: The Past, Present, and Future of the Cancer Registry and Its Efforts to Improve the Quality of Cancer Care," Semin Radiat Oncol, vol. 29, no. 4, pp. 323–325, Oct. 2019, doi: 10.1016/J.SEMRADONC.2019.05.005.

92. M. C. Daly and I. M. Paquette, "Surveillance, Epidemiology, and End Results (SEER) and SEER-Medicare Databases: Use in Clinical Research for Improving Colorectal Cancer Outcomes," Clin Colon Rectal Surg, vol. 32, no. 01, pp. 061–068, 2019.

93. M. J. Landrum and B. L. Kattman, "ClinVar at five years: Delivering on the promise," Hum Mutat, vol. 39, no. 11, pp. 1623–1630, Nov. 2018, doi: 10.1002/HUMU.23641.

94. J. S. Nanda, R. Kumar, and G. P. S. Raghava, "dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes," Sci Rep, vol. 6, Jan. 2016, doi: 10.1038/SREP19340.

95. S. Ullah et al., "The Cancer Research Database (CRDB): Integrated Platform to Gain Statistical Insight into the Correlation between Cancer and COVID-19," JMIR Cancer, vol. 8, no. 2, Apr. 2022, doi: 10.2196/35020.

96. H. Zheng et al., "Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis," Front Oncol, vol. 10, p. 68, Feb. 2020, doi: 10.3389/FONC.2020.00068/BIBTEX.

97. C. P. Goswami and H. Nakshatri, "PROGgeneV2: Enhancements on the existing database," BMC Cancer, vol. 14, no. 1, pp. 1–6, Dec. 2014, doi: 10.1186/1471-2407-14-970/FIGURES/3.

98. R. Kumar et al., "CancerDR: Cancer drug resistance database," Sci Rep, vol. 3, 2013, doi: 10.1038/srep01445.

99. S. H. Liu et al., "DriverDBv3: a multi-omics database for cancer driver gene research," Nucleic Acids Res, vol. 48, no. D1, pp. D863–D870, Jan. 2020, doi: 10.1093/NAR/GKZ964.

100. L. Cheng et al., "LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse," Nucleic Acids Res, vol. 47, no. D1, pp. D140–D144, Jan. 2019, doi: 10.1093/NAR/GKY1051.

101. Y. Gao et al., "Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data," Nucleic Acids Res, vol. 49, no. D1, pp. D1251–D1258, Jan. 2021, doi: 10.1093/NAR/GKAA1006.

102. L. J. Carithers and H. M. Moore, "The Genotype-Tissue Expression (GTEx) Project," https://home.liebertpub.com/bio, vol. 13, no. 5, pp. 307–308, Oct. 2015, doi: 10.1089/BIO.2015.29031.HMM.

103. C. J. Liu, F. F. Hu, M. X. Xia, L. Han, Q. Zhang, and A. Y. Guo, "GSCALite: a web server for gene set cancer analysis," Bioinformatics, vol. 34, no. 21, pp. 3771–3772, Nov. 2018, doi: 10.1093/BIOINFORMATICS/BTY411.

104. A. Hamosh, J. S. Amberger, C. Bocchini, A. F. Scott, and S. A. Rasmussen, "Online Mendelian Inheritance in Man (OMIM®): Victor McKusick's magnum opus," Am J Med Genet A, vol. 185, no. 11, pp. 3259–3265, Nov. 2021, doi: 10.1002/AJMG.A.62407.

105. Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," Nucleic Acids Res, vol. 45, no. W1, pp. W98–W102, Jul. 2017, doi: 10.1093/NAR/GKX247.

106. B. Wen, X. Wang, and B. Zhang, "PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations," Genome Res, vol. 29, pp. 485–493, Jan. 2019.

107. M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data 2016 3:1, vol. 3, no. 1, pp. 1–9, Mar. 2016, doi: 10.1038/sdata.2016.18.

108. A. Pavlopoulou, D. A. Spandidos, and I. Michalopoulos, "Human cancer databases (Review)," Oncology Reports, vol. 33, no. 1. Spandidos Publications, pp. 3–18, Jan. 01, 2015. doi: 10.3892/or.2014.3579.