ICELL JOURNAL | RESEARCH ARTICLE

Vol.1No.1(1)

TGDD: A Genomics-Based Database for Tumor Treatment Drugs

Wen Li a, b, c, 1, Zhining Xiong a, 1, Kangming Hed, Zhining Zhang d, Huilin Wei a, Qiuxia Meng e,Tengyue Yan f, Yuxiao He a, Xuanyu Pan g, Yanling Hu* a, b, c, d

DIO:https://dx.doi.org/10.71373/UYBT8671
Submitted 29 September 2024
Accepted 26 October 2024
Published 1 November 2024

With advancements in cancer genomics, precision medicine is increasingly being applied in cancer treatment. However, the fragmentation of genetic mutation-related information and the lack of standardized interpretation tools present significant challenges to its effective implementation and clinical decision-making. To address these challenges, we have developed the Tumor Genomic Drug Database (TGDD), a comprehensive genomics-based cancer drug treatment resource. It integrates data on genes, molecular characteristics, cancer types, and therapeutic drugs from established cancer knowledge bases such as CIViC, COSMIC, and OncoKB, providing a unified and standardized clinical support tool. Currently, it encompasses 19,557 entries, including 391 gene datasets, 525 cancer types, 5,322 drug combinations, and 8,109 mutation coordinates. This resource assists clinicians in formulating personalized treatment plans and supports cancer research and drug development.

Introduction

Over the past decade, advances in cancer genomics have significantly driven the development of precision medicine [1-4]. As genomic sequencing technologies have matured, prospective clinical sequencing for tumors has become a crucial component of standard cancer care. By detecting gene mutations in a patient's tumor, especially those driver mutations that have a substantial impact on tumor growth and progression, precision medicine can help select the most appropriate targeted therapies, leading to personalized treatment plans that improve therapeutic outcomes [5-8].

Thus, the key to achieving precision oncology lies in accurately identifying and interpreting driver mutations and their implications for treatment strategies and prognosis[9-12]. This process necessitates the integration of extensive literature, guidelines, and expert opinions. However, current efforts often rely on disparate research centers and hospitals, leading to a lack of consistency and standardization. This fragmented approach not only limits a systematic understanding of the relationship between individual genetic variations and precision treatments but can also risks delays and errors in clinical decision-making [13, 14].

To tackle these challenges, there is an urgent need for a standardized clinical support tool that integrates and interprets the relationship between driver mutations and precision therapeutic drugs. Such a tool would consolidate dispersed information into a unified format, empowering clinicians of varying expertise levels to accurately understand genomic variations in patient tumor samples and make optimal treatment decisions. Although existing databases and knowledge bases, such as OncoKB [15] , CIViC [16], COSMIC [17, 18], and ClinVar [19, 20], play important roles in cancer genomics research and clinical applications, they face challenges related to content acquisition, licensing restrictions, and data updates.

This study introduces the Tumor Genomic Drug Database (TGDD), designed to address the fragmentation of cancer genomics data. Integrating information from bases like OncoKB [15], CIViC [16], and COSMIC [17, 18], TGDD organizes data comprehensively across four critical dimensions: genes, molecular features, cancer types, and precision therapeutic drugs. By consolidating and integrating clinically validated gene mutations and treatment efficacy data, TGDD overcomes data integration gaps, ensuring accuracy and completeness. This resource aids clinicians in making informed decisions and offers researchers a powerful tool to advance cancer research and drug development, aiming to optimize treatment outcomes for specific tumor types and driver mutations.

2 Methods

2.1 Data Sources

To ensure comprehensive and diverse data coverage, TGDD adheres to criteria that the selected data must include annotations of genes, molecular features, cancer types, and therapeutic drugs. It primarily sources data from three key databases: CIViC [16], COSMIC [17, 18], and OncoKB [15]. Specifically, CIViC provides clinical evidence for gene mutations and their potential impact on treatment efficacy; COSMIC offers data on actionable and resistance projects related to specific mutations; and OncoKB focuses on clinical evidence for drugs targeting various gene variants, molecular features, and cancer types. To further enhance data completeness, TGDD also incorporates genomic coordinate information for mutations from the ClinVar[19, 20] database.

2.2 Data Acquisition and Processing

The data acquisition process is as follows: First, literature searches are conducted in professional databases, such as PubMed (https://pubmed.ncbi.nlm.nih.gov), to identify the required database resources, following the guidelines provided in the literature. Relevant database websites are then accessed to obtain original data using their download functionalities. Where download options are unavailable, data are collected manually.

During data processing, raw data are initially screened using Excel to eliminate missing entries, retaining only those with information on genes, molecular features, cancer types, and treatments. The filtered data are then organized and stored according to TGDD’s specifications. To ensure completeness and consistency, Python and other tools are employed to merge and deduplicate data from multiple sources. Any data with missing values or anomalies is verified and supplemented from their original databases, and deletions are made when necessary. Finally, the cleaned data are stored in a MySQL database and regularly backed up to ensure security.

To support further research, detailed genomic information is supplemented for more molecular features, including genomic coordinates, variant sequences, and variant types. This information is sourced from original databases or the ClinVar database, facilitating deeper analysis of data potentially related to therapeutic drugs.

2.3 Database Construction and Architecture Design

The TGDD system is developed using a range of technologies to ensure efficiency, flexibility, and ease of maintenance. The core framework is ThinkPHP (https://www.thinkphp.cn/), a lightweight PHP tool that supports rapid application development with well-organized code. For data management, MySQL (https://www.mysql.com/) is employed, providing effective data storage and retrieval.

The system architecture is based on the MVC (Model-View-Controller) design pattern, which separates frontend and backend functionalities, thereby enhancing scalability and efficiency. On the frontend, HTML, CSS, and JavaScript are used in conjunction with Ajax and JSON to enable dynamic data loading and updates, improving user experience. The Layui frontend framework is also integrated to further optimize interface presentation.

Additionally, the SMARTY (https://www.smarty.net/) template engine is utilized to separate page layout from logic, increasing frontend development flexibility. Static caching technology is applied to reduce server load and accelerate page loading speeds, while pseudo-static routing technology is used to improve search engine indexing.

Database Access and Usage

The TGDD database is designed to facilitate medication selection based on the individual molecular characteristics, thereby optimizing drug efficacy and reducing adverse reactions. Users can access the database via the official website (http://tgddb.omics.henbio.com), which offers detailed information and usage instructions. The website features a user-friendly interface with four main sections: Home, Search, Tutorial, and Contact Us. These sections allow users to understand the database content, retrieve the necessary data efficiently, and receive relevant guidance and technical support.

Result

Data Statistics

TAs of now, TGDD contains a total of 19,557 data entries, detailed in Table 1. This dataset includes 391 entries for gene information, 525 entries for tumor types, 5,322 entries for drug combinations, and 2,686 entries for molecular features.

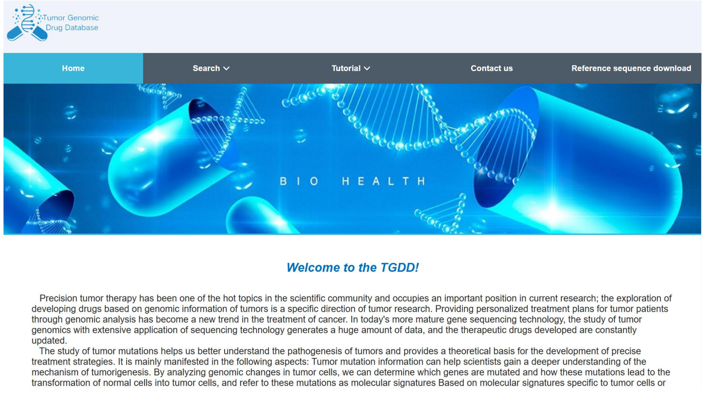
Additionally, there are 8,109 variant coordinate entries and 6,522 variant records linked to molecular features. The current TGDD version annotates 11 types of information. Table 1 indicates that TGDD offers extensive records on drug combination therapy options, and the statistical results provide comprehensive details on molecular features, including gene coordinates and variant records.

Table 1. Comparation of TGDD and Other Database

Item	Quantity		
	TGDD	CIViC	OncoKB
gene	391	611	865
molecular profile	2686	4587	7794
cancer type	525	397	139
drugs	5322	536	139
chromesome NO.	5969	unrecorded	unrecorded
gene coordinates on the chromosome	8109	unrecorded	unrecorded
variant_types	64	3742	unrecorded
HGVS descriptions	6522	unrecorded	none

Home Page Overview

The Home page of the TGDD database is designed with simplicity, providing an overview that helps users quickly grasp the core functionalities of the database (as shown in Figure 1).



Welcome to the TGDD!

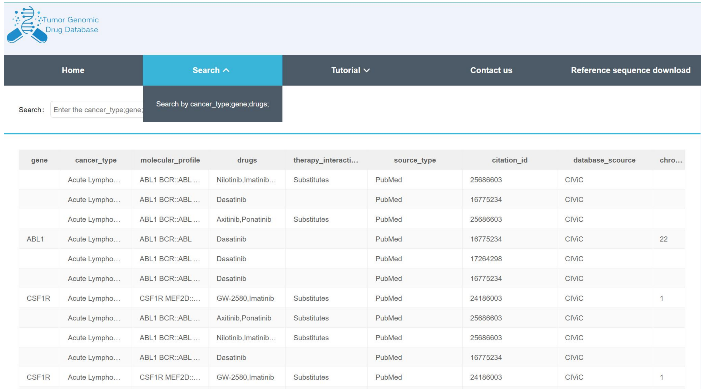
Precision tumor therapy has been one of the hot topics in the scientific community and occupies an important position in current research; the exploration of developing drugs based on genomic information of tumors is a specific direction of tumor research. Providing personalized treatment plans for tumor patients through genomic analysis has become a new trend in the treatment of cancer. In today's more mature gene sequencing technology, the study of tumor genomics with extensive application of sequencing technology generates a huge amount of data, and the therapeutic drugs developed are constantly updated.

The study of tumor mutations helps us better understand the pathogenesis of tumors and provides a theoretical basis for the development of precise treatment strategies. It is mainly manifested in the following aspects: Tumor mutation information can help scientists gain a deeper understanding of the mechanism of tumorigenesis. By analyzing genomic changes in tumor cells, we can determine which genes are mutated and how these mutations lead to the transformation of normal cells into tumor cells, and refer to these mutations as molecular signatures Based on molecular signatures specific to tumor cells or

Figure 1. Screenshot of the Home Page

Search Page Overview and Demonstration

The Search page serves as the core functional module of TGDD, allowing users to retrieve data by entering keywords such as genes, molecular features, tumor types, or drugs. Search results are presented in a detailed tabular format, including various information fields like genes, molecular features, tumor types, drugs, evidence support, database names, chromosome numbers, mutation coordinates, reference sequences, variant types, and HGVS descriptions (as shown in Figure 2).

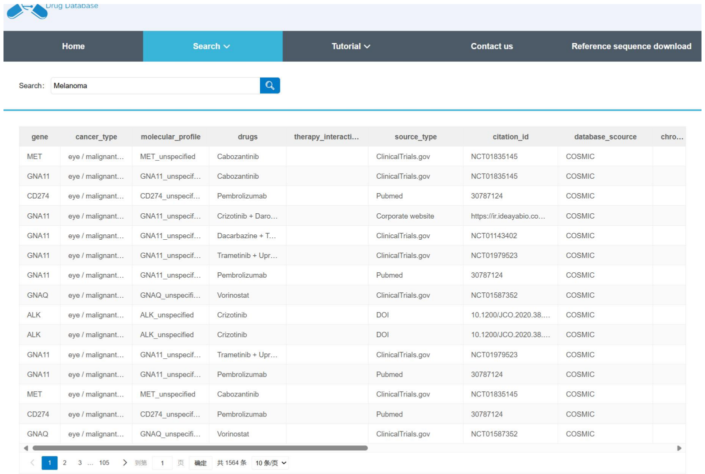

Tumor Genomic
Drug Database
Home
Search
Tutorial
Contact us
Reference sequence download
Search by cancer_type;gene;drugs;

Search: Enter the cancer_type;gene;

gene	cancer_type	molecular_profile	drugs	therapy_interacti...	source_type	citation_id	database_scource	chro...
	Acute Lympho...	ABL1 BCR::ABL ...	Nilotinib,Imatinib...	Substitutes	PubMed	25686603	CIViC	
	Acute Lympho...	ABL1 BCR::ABL ...	Dasatinib		PubMed	16775234	CIViC	
	Acute Lympho...	ABL1 BCR::ABL ...	Axitinib,Ponatinib	Substitutes	PubMed	25686603	CIViC	
ABL1	Acute Lympho...	ABL1 BCR::ABL	Dasatinib		PubMed	16775234	CIViC	22
	Acute Lympho...	ABL1 BCR::ABL ...	Dasatinib		PubMed	17264298	CIViC	
	Acute Lympho...	ABL1 BCR::ABL ...	Dasatinib		PubMed	16775234	CIViC	
CSF1R	Acute Lympho...	CSF1R MEF2D::...	GW-2580,Imatinib	Substitutes	PubMed	24186003	CIViC	1
	Acute Lympho...	ABL1 BCR::ABL ...	Axitinib,Ponatinib	Substitutes	PubMed	25686603	CIViC	
	Acute Lympho...	ABL1 BCR::ABL ...	Nilotinib,Imatinib...	Substitutes	PubMed	25686603	CIViC	
	Acute Lympho...	ABL1 BCR::ABL ...	Dasatinib		PubMed	16775234	CIViC	
CSF1R	Acute Lympho...	CSF1R MEF2D::...	GW-2580,Imatinib	Substitutes	PubMed	24186003	CIViC	1

Figure 2: Screenshot of the Search Page

To illustrate the functionality of the Search page, this article uses Melanoma as an example to demonstrate the search operation. When users enter "Melanoma" into the search box and click the search icon, TGDD returns all data related to Melanoma, presented in a table. This table lists detailed genetic mutation information associated with Melanoma, such as the BRAF V600E mutation detected in a case of Melanoma. This mutation is located at position 140453136 on chromosome 7, affecting the coding of the B-Raf protein, with a variant sequence of NM_004333.4:c.1799T>A (as shown in Figures 3). Further information about this variant can be found in the literature with a PubMed ID of 28891408. If specific mutation coordinates are not available from the source database or Clinvar database, the corresponding cells in the table will be left blank.



Search: Melanoma

gene	cancer_type	molecular_profile	drugs	therapy_interacti...	source_type	citation_id	database_scource	chro...
MET	eye / malignant...	MET_unspecified	Cabozantinib		ClinicalTrials.gov	NCT01835145	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Cabozantinib		ClinicalTrials.gov	NCT01835145	COSMIC	
CD274	eye / malignant...	CD274_unspecif...	Pembrolizumab		Pubmed	30787124	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Crizotinib + Daro...		Corporate website	https://ir.ideayabio.co...	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Dacarbazine + T...		ClinicalTrials.gov	NCT01143402	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Trametinib + Upr...		ClinicalTrials.gov	NCT01979523	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Pembrolizumab		Pubmed	30787124	COSMIC	
GNAQ	eye / malignant...	GNAQ_unspecifi...	Vorinostat		ClinicalTrials.gov	NCT01587352	COSMIC	
ALK	eye / malignant...	ALK_unspecified	Crizotinib		DOI	10.1200/JCO.2020.38....	COSMIC	
ALK	eye / malignant...	ALK_unspecified	Crizotinib		DOI	10.1200/JCO.2020.38....	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Trametinib + Upr...		ClinicalTrials.gov	NCT01979523	COSMIC	
GNA11	eye / malignant...	GNA11_unspecif...	Pembrolizumab		Pubmed	30787124	COSMIC	
MET	eye / malignant...	MET_unspecified	Cabozantinib		ClinicalTrials.gov	NCT01835145	COSMIC	
CD274	eye / malignant...	CD274_unspecif...	Pembrolizumab		Pubmed	30787124	COSMIC	
GNAQ	eye / malignant...	GNAQ_unspecifi...	Vorinostat		ClinicalTrials.gov	NCT01587352	COSMIC	

Figure 3. Screenshot of the page showing the results obtained after searching with the keyword "melanoma."

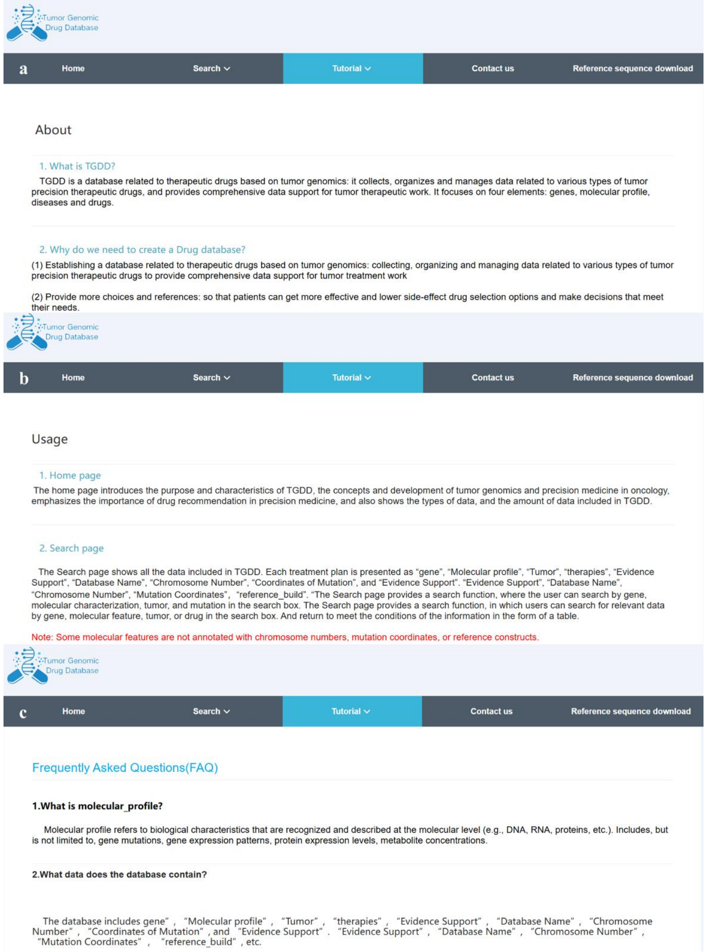
Tutorial Page Overview

To help users effectively utilize TGDD, the Tutorial page offers comprehensive guidance divided into three sections: About, Usage, and FAQ.

About Section: This section provides a detailed introduction to TGDD, including its construction philosophy, research objectives, and data sources. It helps users understand the background and design intentions of the database (as shown in Figure 4(a)).

Usage Section: This section explains the design and usage of various TGDD pages, including Home, Search, Tutorial, and Contact sections. It also outlines the terms of use and key considerations to ensure proper utilization of the database (as shown in Figure 4(b)).

FAQ Section: This section addresses common questions users might encounter, such as how to perform data queries and what information is included in the database. It helps users make better use of TGDD(as shown in Figure 4 (c)).



About

1. What is TGDD?

TGDD is a database related to therapeutic drugs based on tumor genomics: it collects, organizes and manages data related to various types of tumor precision therapeutic drugs, and provides comprehensive data support for tumor therapeutic work. It focuses on four elements: genes, molecular profile, diseases and drugs.

2. Why do we need to create a Drug database?

(1) Establishing a database related to therapeutic drugs based on tumor genomics: collecting, organizing and managing data related to various types of tumor precision therapeutic drugs to provide comprehensive data support for tumor treatment work

(2) Provide more choices and references: so that patients can get more effective and lower side-effect drug selection options and make decisions that meet their needs.

Usage

1. Home page

The home page introduces the purpose and characteristics of TGDD, the concepts and development of tumor genomics and precision medicine in oncology, emphasizes the importance of drug recommendation in precision medicine, and also shows the types of data, and the amount of data included in TGDD.

2. Search page

The Search page shows all the data included in TGDD. Each treatment plan is presented as “gene”, “Molecular profile”, “Tumor”, “therapies”, “Evidence Support”, “Database Name”, “Chromosome Number”, “Coordinates of Mutation”, and “Evidence Support”. “Evidence Support”, “Database Name”, “Chromosome Number”, “Mutation Coordinates”, “reference_build”. “The Search page provides a search function, where the user can search by gene, molecular characterization, tumor, and mutation in the search box. The Search page provides a search function, in which users can search for relevant data by gene, molecular feature, tumor, or drug in the search box. And return to meet the conditions of the information in the form of a table.

Note: Some molecular features are not annotated with chromosome numbers, mutation coordinates, or reference constructs.

Frequently Asked Questions(FAQ)

1.What is molecular_profile?

Molecular profile refers to biological characteristics that are recognized and described at the molecular level (e.g., DNA, RNA, proteins, etc.). Includes, but is not limited to, gene mutations, gene expression patterns, protein expression levels, metabolite concentrations.

2.What data does the database contain?

The database includes gene” , “Molecular profile” , “Tumor” , “therapies” , “Evidence Support” , “Database Name” , “Chromosome Number” , “Coordinates of Mutation” , and “Evidence Support” . “Evidence Support” , “Database Name” , “Chromosome Number” , “Mutation Coordinates” , “reference_build” , etc.

Figure 4. Screenshot of the Tutorial Page. Figures a, b, and c correspond to the About, Usage, and FAQ sections, respectively.

Contact Us Page Overview

TGDD values user feedback and has established a Contact Us page for this purpose. Users can use this page to reach the database administrators via email, allowing them to submit questions, suggestions, or feedback. This submit helps us continually optimize and enhance TGDD’s functionality and services.

Discussion and Future Directions

With the rapid growth of tumor-related data, managing and utilizing this information effectively has become a significant challenge. TGDD addresses this challenge by isolating information on precision treatment plans from numerous complex tumor knowledge bases, providing an efficient query pathway. It offers convenience for users who need quick access to precision treatment plans. Additionally, by integrating data from multiple sources, it offers broader and more comprehensive advantages for drug screening compared to single databases.

Nevertheless, TGDD has limitations. It does not cover all available data, as it focuses on just four key elements, leading to some exclusion of information. Additionally, the database struggles with integrating different naming conventions for the same gene across various sources, which creates inconsistencies (e.g., the same gene segment recorded under different names). Moreover, manual data supplementation risks information loss and adds complexity to ongoing maintenance.

Future work on TGDD will address these issues by: expanding data sources to reduce information silos, optimizing data integration to standardize gene naming, establishing a standardized evaluation system for treatment plans to enhance data reliability, and improving system functionality through regular updates and maintenance. These improvements aim to enhance TGDD's value in precision medicine and advance tumor treatment research.

Reference

1. A. Zehir, R. Benayed, R.H. Shah, A. Syed, S. Middha, et al., Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients, Nature Medicine, 23 (2017) 703-713.
2. J.J. Harding, S. Nandakumar, J. Armenia, D.N. Khalil, M. Albano, et al., Prospective Genotyping of Hepatocellular Carcinoma: Clinical Implications of Next-Generation Sequencing for Matching Patients to Targeted and Immune Therapies, Clinical Cancer Research, 25 (2019) 2116-2126.
3. F. Mosele, J. Remon, J. Mateo, C.B. Westphalen, F. Barlesi, et al., Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group, Annals of Oncology, 31 (2020) 1491-1505.
4. D. Wang, B. Liu, Z. Zhang, Accelerating the understanding of cancer biology through the lens of genomics, Cell, 186 (2023) 1755-1771.
5. R.A.-O. Nussinov, H. Jang, C.J. Tsai, F.A.-O. Cheng, Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers.
6. P.K.-S. Ng, J. Li, K.J. Jeong, S. Shao, H. Chen, et al., Systematic Functional Annotation of Somatic Mutations in Cancer, Cancer Cell, 33 (2018) 450-462.e410.
7. M.H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, et al., Comprehensive Characterization of Cancer Driver Genes and Mutations, Cell, 173 (2018) 371-385.e318.
8. G. Koh, A. Degasperi, X. Zou, S. Momen, S. Nik-Zainal, Mutational signatures: emerging concepts, caveats and clinical applications, Nature Reviews Cancer, 21 (2021) 619-637.
9. M.T. Chang, S. Asthana, S.P. Gao, B.H. Lee, J.S. Chapman, et al., Identifying recurrent mutations in cancer reveals widespreadlineage diversity and mutational specificity, Nature Biotechnology, 34 (2016) 155-163.
10. D. Ostroverkhova, T.M. Przytycka, A.R. Panchenko, Cancer driver mutations: predictions and reality, Trends in Molecular Medicine, 29 (2023) 554-566.
11. S.H. Shin, A.M. Bode, Z. Dong, Addressing the challenges of applying precision oncology, npj Precision Oncology, 1 (2017) 28.
12. A.M. Bode, Z. Dong, Recent advances in precision oncology research, npj Precision Oncology, 2 (2018) 11.
13. A.P.G.C. The, A.P.G.C. The, F. André, M. Arnedos, A.S. Baras, etal.,AACRProjectGENIE:PoweringPrecisionMedicinethroughanInternationalConsortium,CancerDiscovery,7(2017)818 -831.
14.E.Wong,N.Bertin,M.Hebrard,R.Tirado-Magallanes,C.Bellis,etal.,TheSingaporeNationalPrecisionMedicineStrategy, NatureGenetics,55(2023)178-186.
15.oncokbaprecisiononcologyknowledgebasesourcejcopr ecisoncol2017jul2017.pdf>.
16.M.Griffith,N.C.Spies,K.Krysiak,J.F.McMichael,A.C.Coffman,etal。 CIViCisacommunityknowledgebaseforexpertcrowdsourcingtheclinicalinterpretationofvariantsincancer,NatureGenetics,49(2017)170-174.
17.J.G.Tate,S.Bamford,H.C.Jubb,Z.Sondka,D.M.Beare,etal., COSMIC:theCatalogueOfSomaticMutationsInCancer,Nucle icAcidsResearch,47(2019)D941-D947.
18.Z.Sondka,N.B.Dhir,D.Carvalho-Silva,S.Jupe,Madhumita, etal.,COSMIC:acurateddatabaseofsomaticvariantsandclinicaldataforcancer,NucleicAcidsResearch,52(2024)D1210-D1217.
19.M.J.Landrum,S.Chitipiralla,G.R.Brown,C.Chen,B.Gu,et al.,ClinVar:improvementstoaccessingdata,NucleicAcidsResearch,48(2020)D835-D844.
20. M.J. Landrum, J.M. Lee, M. Benson, G.R. Brown, C. Chao, et al., ClinVar: improving access to variant interpretations and supporting evidence, Nucleic Acids Research, 46 (2018) D1062-D1067.

Declaration of Competing interests: The author declares no competing interests.**Fundings:**This work is supported by the National Natural Science Foundation of China (82160537), the Key Research and Development Program of Guangxi(GuikeAB22035027, GuikeAB24010148), and the National Key Research and Development Program of China(2023YFC2605400).**Availability:**The website can be freely accessed at http://tgddb.omics.henbio.com.