TIMEDB

tumor immune micro-environment cell composition database with automatic analysis and interactive visualization

Wang, Xueying; Chen, Lingxi; Liu, Wei; Zhang, Yuanzheng; Liu, Dawei; Zhou, Chenxin; Shi, Shuai; Dong, Jiajie; Lai, Zhengtao; Zhao, Bingran; Zhang, Wenjingyu; Cheng, Haoyue; Li, Shuaicheng

Published in:
Nucleic acids research

Published: 06/01/2023

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Published version (DOI):
10.1093/nar/gkac1006

Publication details:
Wang, X., Chen, L., Liu, W., Zhang, Y., Liu, D., Zhou, C., Shi, S., Dong, J., Lai, Z., Zhao, B., Zhang, W., Cheng, H., & Li, S. (2023). TIMEDB: tumor immune micro-environment cell composition database with automatic analysis and interactive visualization. *Nucleic acids research*, *51*(D1), D1417–D1424. https://doi.org/10.1093/nar/gkac1006

Published online 18 November 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D1417–D1424***
https://doi.org/10.1093/nar/gkac1006

TIMEDB: tumor immune micro-environment cell composition database with automatic analysis and interactive visualization

Xueying Wang [1,2,†], Lingxi Chen [1,2,†], Wei Liu [1,†], Yuanzheng Zhang [3], Dawei Liu [4], Chenxin Zhou [5], Shuai Shi [6], Jiajie Dong [1], Zhengtao Lai [5], Bingran Zhao [5], Wenjingyu Zhang [5], Haoyue Cheng [7] and Shuaicheng Li [1,2,*]

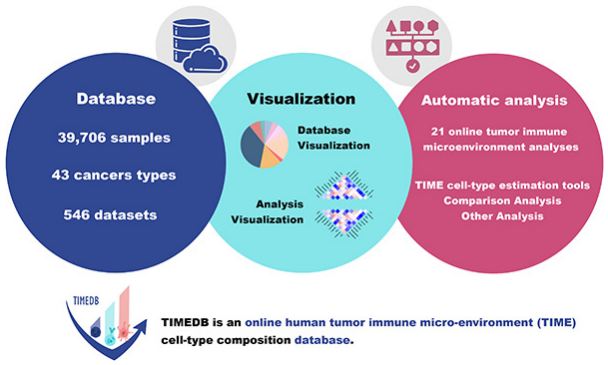
[1]Department of Computer Science, City University of Hong Kong, Hong Kong, China, [2]City University of Hong Kong Shenzhen Research Institute, Guangdong, China, [3]School of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning, China, [4]School of Software, Northeastern University, Shenyang, Liaoning, China, [5]School of Business Administration, Northeastern University, Shenyang, Liaoning, China, [6]School of Creative Media, City University of Hong Kong, Hong Kong, China and [7]Department of Clinical Pathology, Capital Medical University, Beijing, China

Received August 14, 2022; Revised September 30, 2022; Editorial Decision October 15, 2022; Accepted October 21, 2022

ABSTRACT

Deciphering the cell-type composition in the tumor immune microenvironment (TIME) can significantly increase the efficacy of cancer treatment and improve the prognosis of cancer. Such a task has benefited from microarrays and RNA sequencing technologies, which have been widely adopted in cancer studies, resulting in extensive expression profiles with clinical phenotypes across multiple cancers. Current state-of-the-art tools can infer cell-type composition from bulk expression profiles, providing the possibility of investigating the inter-heterogeneity and intra-heterogeneity of TIME across cancer types. Much can be gained from these tools in conjunction with a well-curated database of TIME cell-type composition data, accompanied by the corresponding clinical information. However, currently available databases fall short in data volume, multi-platform dataset integration, and tool integration. In this work, we introduce TIMEDB (https://timedb.deepomics.org), an online database for human tumor immune microenvironment cell-type composition estimated from bulk expression profiles. TIMEDB stores manually curated expression profiles, cell-type composition profiles, and the corresponding clinical information of a total of 39,706 samples from 546 datasets across 43 cancer types. TIMEDB comes readily equipped with online tools for automatic analysis and interactive visualization, and aims to serve the community as a convenient tool for investigating the human tumor microenvironment.

GRAPHICAL ABSTRACT



INTRODUCTION

In cancer, the tumor immune microenvironment (TIME) is characterized by dynamic interactions between tumor and immune cells (1,2). Deciphering the cell-type composition in the TIME can significantly improve cancer prognosis and increase the efficacy of cancer treatment (3). TIME cell-type composition could illuminate details on how tumor cells escape the immune response. In clinical trials, TIME cell-type composition can stratify patients and assign the most appropriate treatment regimen according to the target cell types, thus enhancing the treatment options and ultimately improving overall survival (4).

*To whom correspondence should be addressed. Tel: +852 3442 9412; Email: shuaicli@cityu.edu.hk
†The authors wish it to be known that, in their opinion, the first three authors should be considered Joint First Authors.

Wet-lab approaches like fluorescence-activated cell sorting and immunohistochemical staining have been adopted as the gold standards for assessing the number of immune cells in tumor samples. However, flow cytometry demands plenty of cells, which limits its use in tumor biopsies. Immunohistochemical staining estimates individual tumor dissection slides, which cannot represent the heterogeneous immune panorama of the tumor. Additionally, both methods merely infer a small number of cell types. Recent breakthroughs in single-cell RNA sequencing (scRNA-Seq) have been used to catalog cell type and status. However, it is still costly and tedious for routine clinical applications (5–7).

Microarray and bulk RNA sequencing (RNA-Seq) have been widely adopted in cancer studies, accumulating extensive bulk expression profiles with clinical phenotypes in multiple cancers (6). Present state-of-the-art tools estimate the cell-type composition from bulk expression profiles, allowing the investigation of inter-heterogeneity and intra-heterogeneity of TIME across cancers. However, there are very steep learning curves in running these tools for non-bioinformaticians and in curating the corresponding clinical data for non-clinicians. The task can be made more accessible with a well-curated database that brings together TIME cell-type composition data, the corresponding clinical information, and the tools for effective TIME exploration.

Existing online databases that estimated TIME cell-type composition from bulk expression profiles include CancerImmunityQTL (7), GEPIA2021 (8), TCIA (9), TISMO (10), and TIMER2.0 (11). However, these databases have several limitations. (i) Most of the TIME cell-type composition profiles are estimated from The Cancer Genome Atlas (TCGA) repositories (12); the TIME cell-type composition profiles from the extensive pool of bulk expression outside TCGA are ineffectively exploited. (ii) The TIME cell-type composition profiles from the state-of-the-art tools are unavailable in the current databases, and the computational biases from different approaches are incompletely presented. (iii) Integration analyses of TIME cell-type composition among multiple datasets are required to infer the landscape of tumor-specific TIME signatures (1,2). However, such integration analyses from multiple datasets are either missing from these databases or only provided rudimentarily. (iv) These databases offer little in the way of interactive visualizations that are helpful to scientific explorations (13,14).

To overcome these limitations, we propose TIMEDB (https://timedb.deepomics.org), an online human tumor immune microenvironment cell-type composition database with automatic analysis and interactive visualization. TIMEDB has three key features: (i) A curated TIME database of more than 39,706 samples and 546 datasets across 43 cancers, including bulk expression data, clinical data, TIME cell-type composition data estimated by ten state-of-the-art tools, C1-C6 immune subtyping data, *etc*. (ii) Automatic interfaces of 21 tumor immune microenvironment analyses, including TIME cell-type composition estimation, survival, *etc*. (iii) Interactive visualizations of gene expression, TIME cell-type composition estimation, patient subtyping, survival, and correlation result from the database or analysis.

MATERIALS AND METHODS

Data collection

We first summarized and collected gene expression data for 32 cancer types from Cancer Genome Atlas (TCGA) (12) and six cancer types from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (Supplementary Table S1) (15) . TCGA clinical data and raw RNA-Seq (HTSeq) counts were downloaded through TCGAbiolinks (16). Collapsed files from the publicly available TARGET clinical and level 3 gene expression matrix were downloaded from https://ocg.cancer.gov/programs/target/data-matrix (15). Then, we conducted a systematic search in Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geoprofiles/) (17) and the Array Express portal (https://www.ebi.ac.uk/arrayexpress/) (18) with the names of 38 cancer types and 'Expression Profile' as keywords during June 2021–June 2022. As a result, we collected a total of 39,706 samples from 546 publicly available datasets with curated clinical information and gene expression profiles. The brief descriptions and references of the 546 datasets are listed in Supplementary Table S2.

Clinical data curation

We manually assigned each sample to 38 cancer types according to their primary lesions and metastatic statuses. Tumors other than the 38 types were assigned to SCLC (Small Cell Lung Cancer), CLL (Chronic Lymphocytic Leukemia), RT (Rhabdoid Tumor), CML (Chronic Myelogenic Leukemia), and Other, resulting a total of 43 cancer types except 'Other' (Supplementary Table S1). We curated two primary endpoints for each dataset: overall survival (OS) and progression-free survival (PFS). OS refers to the date of diagnosis to death or the last follow-up, and PFS indicates the date of recurrence/relapse/progression or last follow-up. Tumor subtype, stage, grade, type, metastasis, treatment response, *etc.*, are curated if available. All manually curated profiles are checked by at least three individuals. The curation details can be viewed at the 'Curated Clinical Headers' on each dataset information page. Here, we also included any available normal or blood samples labeled with 'tumor_subtype: normal' in each dataset.

Bulk expression profiles preprocessing

We first transferred the array probe id (microarray) or the gene ensemble id (RNA-Seq) to the HUGO gene symbol for each data set. Duplicated gene entries are aggregated with their mean expression data. The normalization and transformation of the expression data conform to the guiding principles of Quackenbush (19), Ali *et al.* (1), and Gao *et al.* (2). We applied quantile normalization, $\log_2$, and scale to microarray intensity datasets. We converted the raw count of RNA-Seq into logarithmic transcripts per kilobase million ($\log_2$ TPM) values. The preprocessing steps for each dataset are recorded in the 'Dataset' table with the 'TIMEDB_RNA_process' field.

Batch effect elimination

In an analysis involving multiple datasets, we employed sva (20) to eliminate batch effects of expression profiles from different datasets or platforms. We applied Uniform Manifold Approximation and Projection (UMAP) (21) into the expression profiles to check the efficacy of the batch effect elimination. In the UMAP plot (Supplementary Figure S1A), each point represents a sample. Before batch effect removal, samples in different datasets and platforms stay apart in the UMAP plot. After batch-effect removal, samples with the same molecular characteristics from different datasets or platforms tend to be mixed (Supplementary Figure S1B).

TIME cell-type composition quality control, estimation, and other data processing

The details of the TIME cell-type composition quality control, estimation, and further data processing are recorded in Supplementary Methods.

Platform development

TIMEDB is hosted on a CentOS 7.4 server with 128GB memory and 60TB storage. The backend is supported by in-house framework (22), consisting of Ruby (v2.7.1), Ruby on Rails (v6.0.2), Apache (v2.4.6), and PostgreSQL (v12.3). The frontend support includes HTML5, Bootstrap4, ES6, Node.js, Vue.js (v2.6.10), and Oviz (13) (https://oviz.org). We have tested TIMEDB on the following operating systems and browsers. Linux: Firefox, Chrome; macOS: Chrome, Firefox, Safari, Edge; Windows: Firefox, Edge, and Chrome.

RESULTS

TIMEDB database

TIMEDB holds curated gene expression profiles and the corresponding clinical information (basic characteristics, survival, *etc.*) from a total of 39,706 samples across 546 datasets across 43 cancer types (Figure 1, Supplementary Table S1, and Supplementary Figures S2–S3). Sequencing protocols include microarray and RNA-Seq. The data sources consist of TCGA, TARGET, GEO (17), and ArrayExpress (18).

TIMEDB stores the TIME cell-type composition profiles estimated by ten state-of-the-art methods. TIME cell-type composition estimation methods are classified into regression-based deconvolution approaches and gene set enrichment-based approaches. ABIS performs deconvolution of the absolute cell proportion of 29 immune cell types on RNA-Seq and microarray data (23). CIBERSORT infers the fraction of 22 TIME cell types from gene expression data (24). This method relies on cell-specific gene signatures and Nu-support vector regression. CIBERSORTx, which extends on CIBERSORT, incorporates ComBat to address possible cross-platform batch effects in gene exerssion data (25). EPIC quantitatively analyzes the ratio of immune and non-tumor cells in TIME from expression profiles by applying constrained least squares regression (26). The quanTIseq method quantifies the proportion of ten different immune cell types and the proportion of other noncharacteristic cells from bulk gene expression of heterogeneous samples. It solves cell-type composition devolution problems by constrained least squares regression (27). TIMER selects marker genes negatively correlated with tumor purity for 32 cancer types from TCGA. It adopts constrained least squares to infer the fraction of six immune cell types (28). TIMEDB database stores the predicted cell-type composition profiles of ABIS, CIBERSORT, CIBERSORTx, EPIC, quanTIseq, and TIMER of the datasets 513, 513, 507, 37, 512, and 455, respectively.

In addition, there are four enrichment-based tools. ConsensusTME (29) is an ssGSEA-based approach. It first integrates gene sets from Bindea *et al.* (30), Davoli *et al.* (31), Danaher *et al.* (32), CIBERSORT LM22 (24), MCPcounter (33), TIMER (28), and xCell (34). Then, consensus marker genes for 18 cell types and 32 TCGA cancer types are curated. This package allows users to choose a type of TCGA cancer and uses ssGSEA to generate normalized enrichment scores and infer the relative TIME cell-type fractions. ImmuCellAI is an online cell abundance identifier based on ssGSEA, explicitly focusing on determining the abundance of 18 T cell subtypes (35). MCPcounter quantifies the absolute abundance of eight immune cell types and two stromal cell types from gene expression data via ssGSEA (33). xCell is a cell quantitative analysis tool based on ssGSEA; it calculates the enrichment fraction of 64 types of immune and nonimmune cells (34). In the TIMEDB database, we obtained ConsensusTME-, ImmuCellAI-, MCPcounter-, and xCell-inferred cell abundance for the 452, 475, 483, and 461 datasets, respectively.

Furthermore, TIMEDB provides the consensus and integrated TIME cell-type composition estimation by ten state-of-the-art tools for each dataset. The consensus result removes cell types that occur only once among ten tools. The integration result means that the cell-type composition is inferred from all tools, and the composition profiles of all cell types are combined into one file.

TIMEDB also keeps *immune subtypes* predicted by ImmuneSubtypeClassifier (36,37), which forms a part of the 'Immune Landscape of Cancer' (38). The tool accepts RNA sequencing data as input and calculates the probabilities for a patient concerning six immune statuses, including C1 (wound healing), C2 (IFN-γ dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGF-β dominant). Currently, TIMEDB contains the ImmuneSubtypeClassifier-inferred C1–C6 subtyping results for 511 curated datasets.

The 'Immune Landscape of Cancer' (38) also asserts seventy-nine immunomodulators. These immunomodulating proteins can be categorized into TNF, MHC Class II, Immunoglobulin, or CXC chemokine gene family; they can also be classified into ligand, receptor, or antigen presentation. From an immune checkpoint view, they can be inhibitory or stimulatory. In the TIMEDB database, we provide 520 immunomodulatory gene expression profiles from curated datasets.

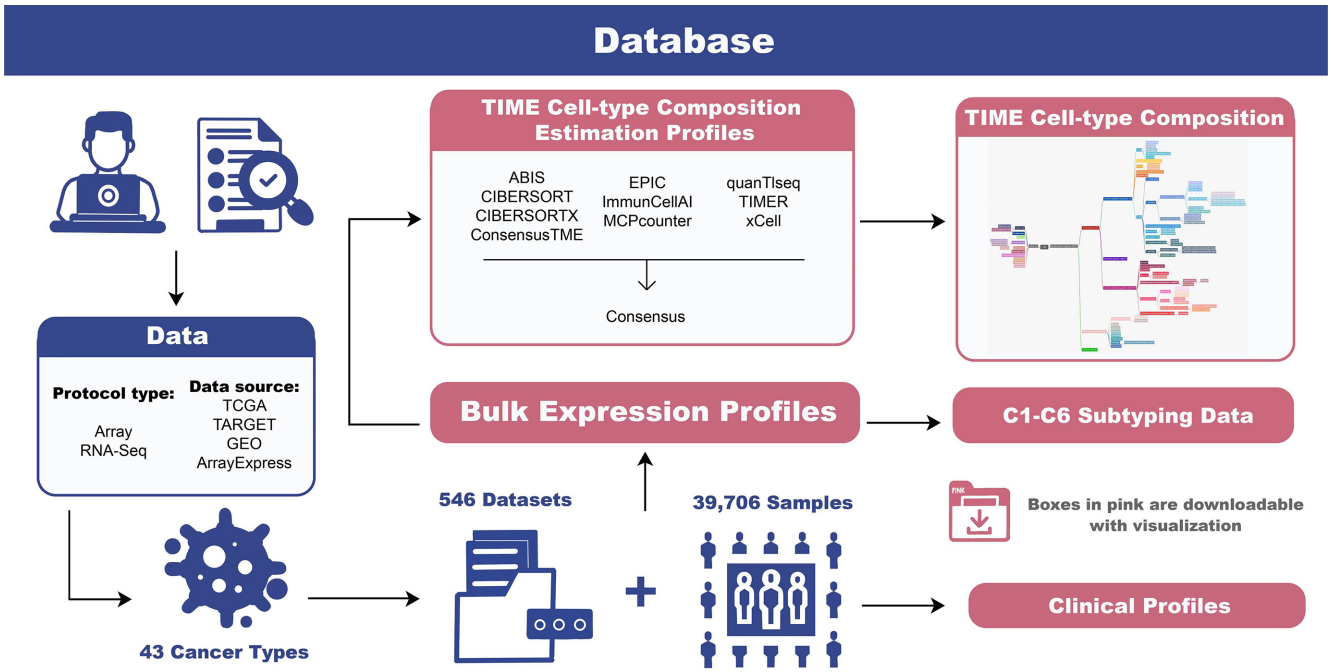


Figure 1. The overview of TIMEDB database design.

TIMEDB analysis

TIMEDB offers an automatic TIME analysis workflow for researchers to run their newly sequenced bulk expression profiles (Figure 2). This workflow includes all the tools mentioned earlier for TIME cell-type composition estimation, TIME cell-type composition-guided patient subtyping, survival analysis, correlation analysis, differential expression analysis, *etc*. Users can upload their gene expression data or select TIMEDB datasets for independent or integrated analyzes with simple mouse operations.

Regarding TIME estimation, with the exception of CIBERSORTx and ImmunecellAI, which only provide online interfaces, all remaining composition estimation tools are included in the TIMEDB analysis interfaces. We have five deconvolution-based interfaces: 'TIMEDB Deconv ABIS', 'TIMEDB Deconv CIBERSORT', 'TIMEDB Deconv EPIC', 'TIMEDB Deconv quanTIseq', and 'TIMEDB Deconv TIMER'. Similarly, we have three enrichment-based analysis interfaces: 'TIMEDB Deconv ConsensusTME', 'TIMEDB Deconv MCPcounter', and 'TIMEDB Deconv xCell'. TIMEDB also includes the LinSeed software into the analysis interface 'TIMEDB Deconv LinSeed'. Unlike supervised prediction, Linseed utilizes nonnegative matrix factorization to reveal the TIME cell-type composition without prior knowledge of the cell types and their abundances (39). It predicts the fraction of *k* TIME cell subtypes and the gene signatures of the corresponding cell subtypes. Users need to manually annotate resolved cell subtypes.RNA-Seq

Furthermore, TIMEDB implements 'TIME Consensus RNA-Seq', 'TIME Consensus Array', 'TIMEDB ALL RNA-Seq', and 'TIMEDB ALL Array' to provide comparison analysis among different datasets and estimation tools. 'Consensus' mode keeps cell types occurring more than twice among different tools, and the 'All' mode combines absolute results for all TIME cell types.

TIMEDB also supports several downstream immune analyses. Molecular subtyping of tumors is an approach to categorizing cancer patients into different subgroups based on molecular profiles and unsupervised clustering tools. ImmuneSubtypeClassifier is included in TIMEDB as the analysis interface 'TIMEDB C1–C6 Subtyping' for researchers to predict the six immune statuses, C1–C6 as explained in database section. Moreover, we provide the TIME cell-type composition-guided patient subtyping. Unsupervised clustering approaches, including *k*-means, hierarchical clustering, and nonnegative matrix factorization, are available in the analysis interface 'TIMEDB Cell Fraction Subtyping'. Furthermore, we group the patients into quantiles based on the TIME cell-type composition and consider the quantiles as the patient subtype. For example, patients can be classified into Q1 B cell (0–25%), Q2 B cell (25–50%), Q3 B cell (50–75%), and Q4 B cell (75–100%) according to the quantiles of the fraction of B cell types.

Survival analysis is essential to draw the clinical implication from TIME cell-type composition. The Kaplan–Meier (KM) study shows the probability of OS or PFS for a patient subgroup at a specific time, illustrating the difference in prognostic events between different subtypes of patients. The hazard ratio (HR) and the odds ratio (OR) of prognostic associations between the subtypes of patients and survival outcomes customized (OS or PFS) are also vital prognostic indications. Here, the 'TIMEDB KM Estimator' and 'TIMEDB HR OR' allow researchers to run their newly sequenced data and custom clinical subgroup labels.

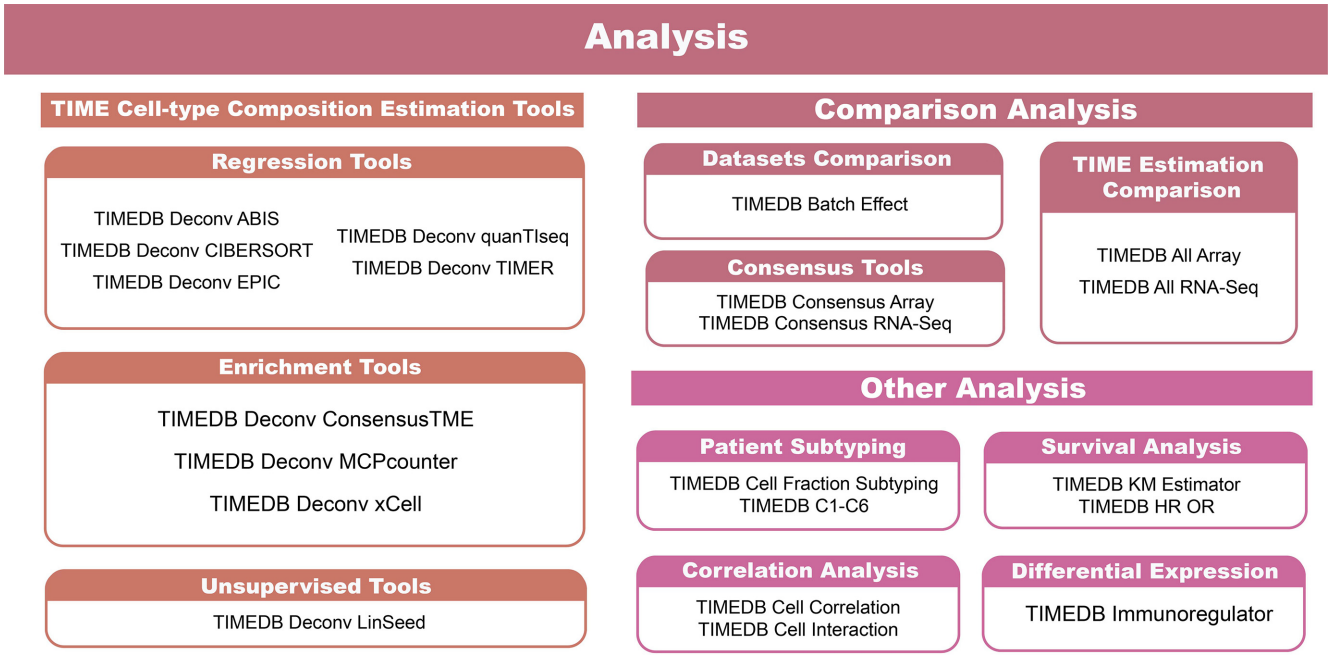


Figure 2. The overview of TIMEDB analysis design.

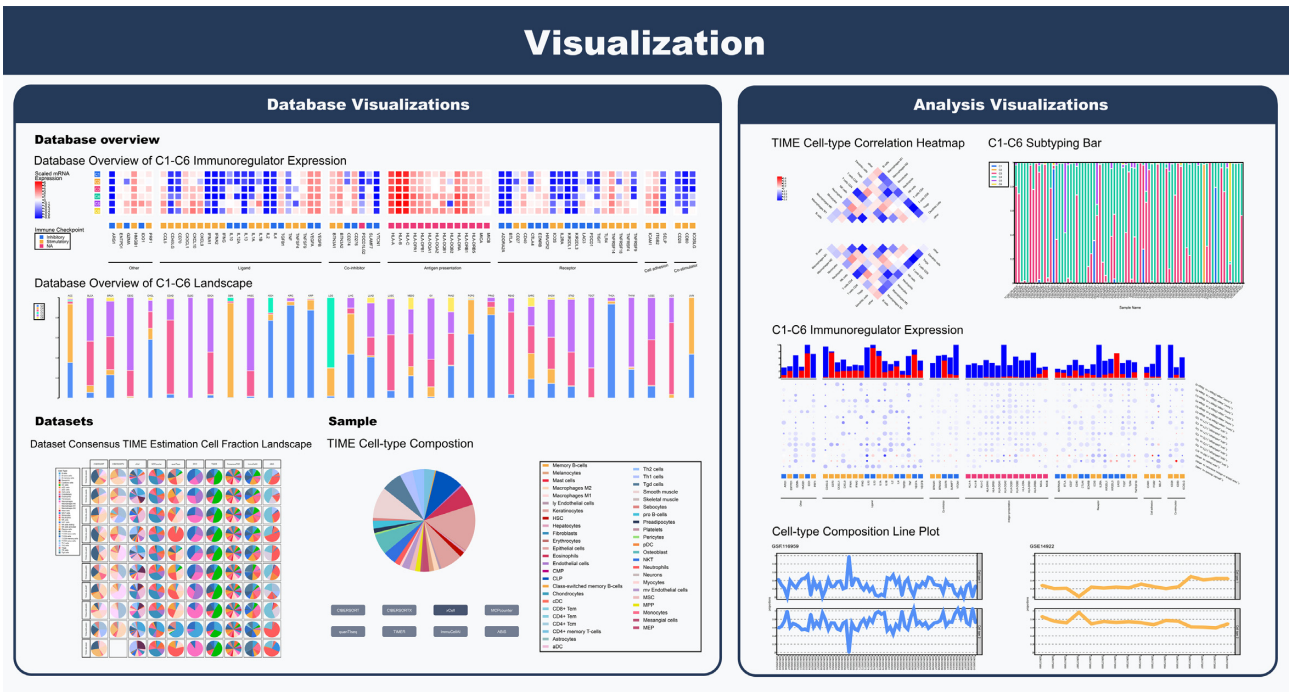


Figure 3. The overview of TIMEDB visualization design.

Cell-type interaction and correlation are essential in TIME. We provide 'TIMEDB Correlation' and 'TIMEDB Interaction' for researchers to calculate the Pearson correlation and associated significance test.

Understanding the expressions of immunomodulatory genes in different stages of TIME is critical for immunotherapy (38). TIMEDB provides an analysis interface 'TIMEDB Immunoregulator' that shows the differential expression of immunomodulatory genes among different patient subtypes, including tumor grades, tumor stages, tumor types, *etc.*, wherever available.

Interactive visualization

TIMEDB supports interactive visualizations of TIME cell-type composition estimation (Supplementary Fig-

Table 1. Comparison of existing TIME databases and webservers. '✓' for available and '-' for not applicable. 'x' signifies offered but unworkable. For instance, CancerImmunityQTL offered downloadable datasets, but no dataset could be downloaded when the button was clicked. In addition, CancerImmunityQTL provides some visualizations about the cell-type composition and KM survival estimation, but the current version does not successfully display these visualizations. 'o' means only a few are included. For example, TCIA merely collects one GEO dataset GSE78220. KM: Kaplan Meier.

			Human TIME database			Human TIME analysis webserver		Mouse TIME database
Functionality			TIMEDB (2022)	CancerImmunityQTL (2021)	TCIA (2016)	GEPIA (2021)	TIMER2.0 (2020)	TISMO (2022)
Database	Data source	TCGA	✓	✓	✓	✓	✓	-
		TARGET	✓	-	-	-	-	-
		ArrayExpress	✓	-	-	-	-	-
		GEO	✓	-	○	-	-	-
	TIME	Bulk expression profiles	✓	x	✓	-	-	-
	basic	Curated clinical profiles	✓	-	✓	-	-	-
	TIME	ABIS	✓	-	-	-	-	-
	cell-type	CIBERSORT	✓	x	✓	-	-	-
	composition	CIBERSORTx	✓	-	-	-	-	-
		ConsensusTME	✓	-	-	-	-	-
		EPIC	✓	-	-	-	-	-
		ImmucellAI	✓	-	-	-	-	-
		MCPcounter	✓	-	-	-	-	-
		quanTIseq	✓	-	-	-	-	-
		TIMER	✓	-	-	-	-	-
		xCell	✓	-	-	-	-	-
		Consensus	✓	-	-	-	-	-
		ALL	✓	-	-	-	-	-
	TIME	Patients subtyping	✓	x	✓	-	-	-
	advanced	Survival	✓	x	✓	-	-	-
		Immunoregulator	✓	-	-	-	-	-
	Total	Cancer types	43	33	20	-	-	-
	number	Datasets number	546	33	24	-	-	-
		Sample number	39706	8514	9672	-	-	-
Analysis	TIME	ABIS	✓	-	-	-	-	-
	cell-type	CIBERSORT	✓	-	-	✓	✓	✓
	composition	ConsensusTME	✓	-	-	✓	✓	✓
		EPIC	✓	-	-	-	-	-
		Linseed	✓	-	-	-	-	-
		MCPcounter	✓	-	-	✓	✓	✓
		quanTIseq	✓	-	-	-	✓	✓
		TIMER	✓	-	-	-	✓	✓
		xCell	✓	-	-	-	✓	✓
		Consensus	✓	-	-	-	✓	✓
		ALL	✓	-	-	-	✓	✓
		Multiple datasets	✓	-	-	-	✓	✓
	Others	Patients subtyping	✓	-	-	-	-	-
		KM estimation	✓	-	-	✓	-	-
		Hazard ratio	✓	-	-	-	-	-
		Odds ratio	✓	-	-	-	-	-
		Cell correlation	✓	-	-	✓	-	-
		Cell interaction	✓	-	-	-	-	-
		Immunoregulator	✓	-	-	-	-	-
Visualization	Interactive	Informative tooltips	✓	✓	✓	✓	-	-
		Changeable plot size	✓	-	-	-	-	-
		Editable label	✓	-	-	-	-	-
		Color picker	✓	-	-	-	-	-
	Downloadable		✓	x	✓	✓	✓	✓

ures S4 and S5), patient subtyping (Supplementary Figures S6 and S7), survival (Supplementary Figures S8–S10), cell-type correlation, interaction (Supplementary Figures S11 and S12), and immunoregulator data (Supplementary Figure S13) from databases or analyses (Figure 3). Multiple interaction options are available, including informative tooltips, changeable plot size, edited labels, and color picker, *etc.* All visualizations are downloadable in high-quality publication-ready format.

DISCUSSION

Understanding the cell-type composition in the tumor immune microenvironment (TIME) can significantly improve the cancer prognosis and increase cancer treatment efficacy. Here, we present TIMEDB, an online human tumor immune microenvironment cell-type composition database with automatic analysis and interactive visualization. TIMEDB has the following merits compared to existing online TIME cell-type composition databases and webservers (Table 1):

First, TIMEDB manually curated bulk expression and clinical profiles of 39,678 samples and 545 datasets across 43 types of cancer from TCGA, TARGET, GEO, and Array Express. Other databases such as CancerImmunityQTL, GEPIA2021, TCIA and TIMER2.0 only contain 33, 33, 33 and 32 TCGA datasets, respectively. TIMEDB provides the opportunity to effectively explore the heterogeneity of TIME from the extensive pool of bulk expression profiles outside TCGA.

Second, we conducted TIME cell estimation using ten state-of-the-art tools, including ABIS, CIBERSORT, CIBERSORTx, EPIC, quanTIseq, TIMER, ConsensusTME, ImmuneCellAI, MCPcounter, and xCell. Such comprehensive TIME cell-type composition profiles from state-of-the-art tools are unavailable in existing databases. For instance, CancerImmunityQTL, GEPIA2021, TCIA, and TIMER2.0 only adopted one (CIBERSORT), one (CIBERSORT), and three (CIBERSORT, EPIC, quanTIseq), and six (CIBERSORT, EPIC, quanTIseq, TIMER, MCPcounter, xCell) estimation tools, respectively. This comprehensiveness can help avoid possible computational biases in the individual tool. Third, TIMEDB infers the landscape of tumor-specific TIME signatures through integration analysis of TIME cell-type composition among multiple datasets, including the above TIME cell-type composition estimation tools, TIME cell-type composition guided patient subtyping, survival analysis, correlation analysis, differential expression analysis, *etc*. In contrast, CancerImmunityQTL and TCIA databases do not support online analysis. GEPIA2021, TISMO, and TIMER2.0 embed a few TIME cell-type composition tools and do not have inconvenient integration analyses for multiple datasets.

Fourth, TIMEDB provides informative visualizations to boost productive scientific explorations, including gene expression, TIME cell-type composition, patient subtyping, survival, and correlation data from databases or analyses. Multiple interaction options are available, including informative tooltips, changeable plot size, edited labels, optional color picker, *etc*. All visualizations are available for download in a high-quality publication-ready format. In contrast, CancerImmunityQTL, TISMO, and TIMER2.0 only provide static low-resolution pictures.

One thing to note here is that TIMEDB focuses on the cell-type composition profiles estimated from the bulk expression. Recent breakthroughs in scRNA-Seq have been used to classify cell types and states in TIME with single-cell resolution. Currently, we are collecting and collating cell-type composition profiles from scRNA-Seq for the upgraded version of TIMEDB.

With comprehensive TIME cell-type composition profiles, analyses, and visualizations at the pan-cancer level, we envision TIMEDB to become a convenient tool for scientists to investigate the biological understanding of human tumor microenvironment study.

DATA AVAILABILITY

All the data and visualizations described are freely available at https://timedb.deepomics.org. Users can report any issues in https://github.com/deepomicslab/TIMEDB.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mr. Bowen Tan, Miss. Yanfei Wang, Miss. Lijia Che, and Mr. Shuo Yang for providing backend technical support. The authors thank Dr. Yinan Wang and Miss. Mengyao Wang for valuable suggestions. The authors thank Mr. Changzheng Wang and Miss. Jing Li for the discussion. We would like to sincerely thank Dr. Yen Kaow Ng for language refinement.

Author contributions: S.C.L. supervised the project. X.Y.W. and L.X.C. designed the study and implemented the analysis. W.L., Y.Z.Z., D.W.L., and J.J.D. developed the platform. X.Y.W., L.X.C., Y.Z.Z., C.X.Z., Z.T.L., W.L., B.R.Z., W.J.Y.Z., and H.Y.C. collected and curated datasets. S.S. designed the webpage user interface. L.X.C. and X.Y.W. wrote the manuscript. S.C.L. revised the manuscript. All authors read and approved the final manuscript.

FUNDING

CityU/UGC Research Matching Grant Scheme [9229013] and SIRG [CityU SIRG 7020005] Funding for open access charge: CityU/UGC Research Matching Grant Scheme [9229013] and The Science Technology and Innovation Committee of Shenzhen Municipality (Project No. JCYJ20200109143216036).

Conflict of interest statement. None declared.

REFERENCES

1. Ali,H.R., Chlon,L., Pharoah,P.D., Markowetz,F. and Caldas,C. (2016) Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS Med.*, **13**, e1002194.
2. Gao,Y., Chen,L., Cai,G., Xiong,X., Wu,Y., Ma,D., Li,S.C. and Gao,Q. (2020) Heterogeneity of immune microenvironment in ovarian cancer and its clinical significance: a retrospective study. *Oncoimmunology*, **9**, 1760067.
3. Galli,F., Aguilera,J.V., Palermo,B., Markovic,S.N., Nisticò,P. and Signore,A. (2020) Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy. *J. Exp. Clin. Cancer Res.*, **39**, 89.
4. Quezada,S.A., Peggs,K.S., Simpson,T.R. and Allison,J.P. (2011) Shifting the equilibrium in cancer immunoediting: from tumor tolerance to eradication. *Immunol. Rev.*, **241**, 104–118.
5. Galluzzi,L., Senovilla,L., Zitvogel,L. and Kroemer,G. (2012) The secret ally: immunostimulation by anticancer drugs. *Nat. Rev. Drug Disc.*, **11**, 215–33.
6. Sturm,G., Finotello,F., Petitprez,F., Zhang,J.D., Baumbach,J., Fridman,W.H., List,M. and Aneichyk,T. (2019) Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*, **35**, i436–i445.
7. Ru,B., Wong,C.N., Tong,Y., Zhong,J.Y., Zhong,S.S.W., Wu,W.C., Chu,K.C., Wong,C.Y., Lau,C.Y., Chen,I. *et al.* (2019) TISIDB: an integrated repository portal for tumor–immune system interactions. *Bioinformatics*, **35**, 4200–4202.
8. Li,C., Tang,Z., Zhang,W., Ye,Z. and Liu,F. (2021) GEPIA2021: integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.*, **49**, W242–W246.
9. Charoentong,P., Finotello,F., Angelova,M., Mayer,C., Efremova,M., Rieder,D., Hackl,H. and Trajanoski,Z. (2017) Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.*, **18**, 248–262.

10. Zeng,Z., Wong,C.J., Yang,L., Ouardaoui,N., Li,D., Zhang,W., Gu,S., Zhang,Y., Liu,Y., Wang,X. *et al.* (2022) TISMO: syngeneic mouse tumor database to model tumor immunity and immunotherapy response. *Nucleic Acids Res.*, **50**, D1391–D1397.
11. Li,T., Fu,J., Zeng,Z., Cohen,D., Li,J., Chen,Q., Li,B. and Liu,X.S. (2020) TIMER2. 0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.*, **48**, W509–W514.
12. Cancer,Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
13. Jia,W., Li,H., Li,S., Chen,L. and Li,S.C. (2020) Oviz-Bio: a web-based platform for interactive cancer genomics data visualization. *Nucleic Acids Res.*, **48**, W415–W426.
14. Chen,L., Qing,Y., Li,R., Li,C., Li,H., Feng,X. and Li,S.C. (2022) Somatic variant analysis suite: copy number variation clonal visualization online platform for large-scale single-cell genomics. *Brief. Bioinform.*, **23**, bbab452.
15. Farrar,J.E., Schuback,H.L., Ries,R.E., Wai,D., Hampton,O.A., Trevino,L.R., Alonzo,T.A., Guidry Auvil,J.M., Davidsen,T.M., Gesuwan,P. *et al.* (2016) Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. *Cancer Res.*, **76**, 2197–2205.
16. Colaprico,A., Silva,T.C., Olsen,C., Garofano,L., Cava,C., Garolini,D., Sabedot,T.S., Malta,T.M., Pagnotta,S.M., Castiglioni,I. *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.*, **44**, e71.
17. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
18. Parkinson,H., Kapushesky,M., Shojatalab,M., Abeygunawardena,N., Coulson,R., Farne,A., Holloway,E., Kolesnykov,N., Lilja,P., Lukk,M. *et al.* (2007) ArrayExpress–a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.*, **35**, D747–D750.
19. Quackenbush,J. (2002) Microarray data normalization and transformation. *Nat. Genetics*, **32**, 496–501.
20. Leek,J.T., Johnson,W.E., Parker,H.S., Jaffe,A.E. and Storey,J.D. (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, **28**, 882–883.
21. Becht,E., McInnes,L., Healy,J., Dutertre,C.-A., Kwok,I.W., Ng,L.G., Ginhoux,F. and Newell,E.W. (2019) Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotech.*, **37**, 38–44.
22. Jiang,Y., Wang,Y., Che,L., Zhou,Q. and Li,S.C. (2022) GutMeta: online microbiome analysis and interactive visualization with build-in curated human gut microbiome database. bioRxiv doi: https://doi.org/10.1101/2022.09.26.509484, 27 September 2022, preprint: not peer reviewed.
23. Monaco,G., Lee,B., Xu,W., Mustafah,S., Hwang,Y.Y., Carré,C., Burdin,N., Visan,L., Ceccarelli,M., Poidinger,M. *et al.* (2019) RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.*, **26**, 1627–1640.
24. Newman,A.M., Liu,C.L., Green,M.R., Gentles,A.J., Feng,W., Xu,Y., Hoang,C.D., Diehn,M. and Alizadeh,A.A. (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*, **12**, 453–457.
25. Newman,A.M., Steen,C.B., Liu,C.L., Gentles,A.J., Chaudhuri,A.A., Scherer,F., Khodadoust,M.S., Esfahani,M.S., Luca,B.A. and Steiner,D. (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotech.*, **37**, 773.
26. Racle,J., de Jonge,K., Baumgaertner,P., Speiser,D.E. and Gfeller,D. (2017) Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *elife*, **6**, e26476.
27. Finotello,F., Mayer,C., Plattner,C., Laschober,G., Rieder,D., Hackl,H., Krogsdam,A., Loncova,Z., Posch,W., Wilflingseder,D. *et al.* (2019) Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.*, **11**, 34.
28. Li,T., Fan,J., Wang,B., Traugh,N., Chen,Q., Liu,J.S., Li,B. and Liu,X.S. (2017) TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.*, **77**, e108.
29. Jiménez-Sánchez,A., Cast,O. and Miller,M.L. (2019) Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res.*, **79**, 6238–6246.
30. Bindea,G., Mlecnik,B., Tosolini,M., Kirilovsky,A., Waldner,M., Obenauf,A.C., Angell,H., Fredriksen,T., Lafontaine,L., Berger,A. *et al.* (2013) Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*, **39**, 782–795.
31. Davoli,T., Uno,H., Wooten,E.C. and Elledge,S.J. (2017) Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, **355**, eaaf8399.
32. Danaher,P., Warren,S., Dennis,L., D'Amico,L., White,A., Disis,M.L., Geller,M.A., Odunsi,K., Beechem,J. and Fling,S.P. (2017) Gene expression markers of tumor infiltrating leukocytes. *J. Immun. Cancer*, **5**, 18.
33. Becht,E., Giraldo,N.A., Lacroix,L., Buttard,B., Elarouci,N., Petitprez,F., Selves,J., Laurent-Puig,P., Sautès-Fridman,C., Fridman,W.H. *et al.* (2016) Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.*, **17**, 218.
34. Aran,D., Hu,Z. and Butte,A.J. (2017) xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.*, **18**, 220.
35. Miao,Y.-R., Zhang,Q., Lei,Q., Luo,M., Xie,G.-Y., Wang,H. and Guo,A.-Y. (2020) ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.*, **7**, 1902880.
36. Gibbs,D.L. (2020) Robust classification of Immune Subtypes in Cancer. bioRxiv doi: https://doi.org/10.1101/2020.01.17.910950, 18 January 2020, preprint: not peer reviewed.
37. Eddy,J.A., Thorsson,V., Lamb,A.E., Gibbs,D.L., Heimann,C., Yu,J.X., Chung,V., Chae,Y., Dang,K., Vincent,B.G. *et al.* (2020) CRI iAtlas: an interactive portal for immuno-oncology research. *F1000Research*, **9**, 1082.
38. Thorsson,V., Gibbs,D.L., Brown,S.D., Wolf,D., Bortone,D.S., Yang,T.-H.O., Porta-Pardo,E., Gao,G.F., Plaisier,C.L., Eddy,J.A. *et al.* (2019) The immune landscape of cancer. *Immunity*, **51**, 411–412.
39. Zaitsev,K., Bambouskova,M., Swain,A. and Artyomov,M.N. (2019) Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. *Nat. Commun.*, **10**, 2209.