ExpressVis: a biologist-oriented interactive web server for exploring multi-omics data

Xian Liu [1,2,†], Kaikun Xu[1,†], Xin Tao[1,4], Ronghua Yin[1,3], Guangming Ren[1,3], Miao Yu[1,3], Changyan Li[1,3], Hui Chen[1,3], Ke Zhao[1,3], Shensi Xiang[1,3], Huiying Gao[1,3], Xiaochen Bo [2,3,*], Cheng Chang [1,5,*] and Xiaoming Yang[1,3,*]

[1]State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China, [2]Institute of Health Service and Transfusion Medicine, Beijing 100850, China, [3]Beijing Institute of Radiation Medicine, Beijing 100850, China, [4]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China and [5]Research Unit of Proteomics Driven Cancer Precision Medicine, Chinese Academy of Medical Sciences, Beijing 102206, China

ABSTRACT

In the era of life-omics, huge amounts of multi-omics data have been generated and widely used in biomedical research. It is challenging for biologists with limited programming skills to obtain biological insights from multi-omics data. Thus, a biologist-oriented platform containing visualization functions is needed to make complex omics data digestible. Here, we propose an easy-to-use, interactive web server named ExpressVis. In ExpressVis, users can prepare datasets; perform differential expression analysis, clustering analysis, and survival analysis; and integrate expression data with protein–protein interaction networks and pathway maps. These analyses are organized into six modules. Users can use each module independently or use several modules interactively. ExpressVis displays analysis results in interactive figures and tables, and provides comprehensive interactive operations in each figure and table, between figures or tables in each module, and among different modules. It is freely accessible at https://omicsmining.ncpsb.org.cn/ExpressVis and does not require login. To test the performance of ExpressVis for multi-omics studies of clinical cohorts, we re-analyzed a published hepatocellular carcinoma dataset and reproduced their main findings, suggesting that ExpressVis is convenient enough to analyze multi-omics data. Based on its complete analysis processes and unique interactive operations, ExpressVis provides an easy-to-use solution for exploring multi-omics data.

GRAPHICAL ABSTRACT
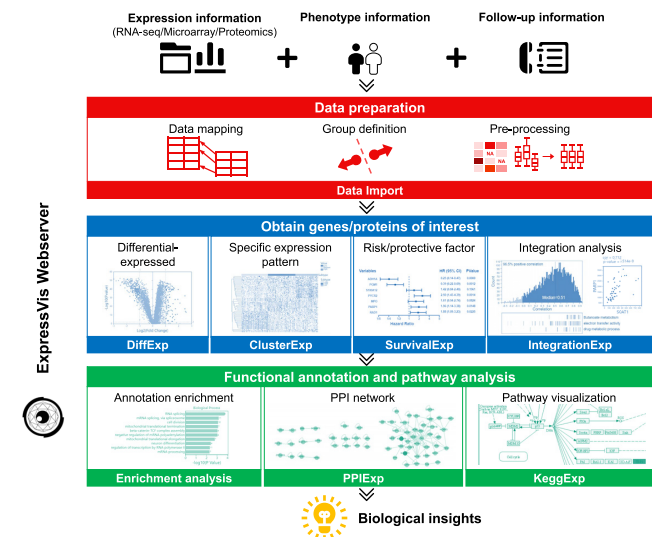
INTRODUCTION

In the era of multi-omics in life science, profiling gene/protein expressions across different conditions has become a common method to explore complex life systems. The rapid development of high-throughput technologies including DNA Microarray, RNA deep sequencing (RNA-seq) (1), and mass spectrometry (2)

has made expression profile data accessible to almost all biomedical researchers. However, extracting biological insights from multi-omics data in expression matrix format remains a huge challenge owing to the high complexity of omics data. Analyzing omics data usually involves multiple steps, including differential expression analysis, clustering analysis, functional enrichment analysis, and integration with protein–protein interaction (PPI) networks and pathway maps. These steps require both data analysis skills and biological knowledge. One of the most popular ways to address this challenge is to develop biologist-oriented platforms that can enable biologists with background knowledge to participate in analyzing and exploring multi-omics data in an iterative exploration process (3,4). These platforms should be biologist-friendly and contain effective visualization functions to make complex omics data digestible for humans (5–7).

Researchers have developed many biologist-oriented platforms to analyze and explore expression profile data. PANDA-view (8) is an easy-to-use, stand-alone tool to analyze and visualize -omics data based on R. However, it only runs on Windows and requires installing a dozen R packages initially. OmicsAnalyst (9) is a powerful web-based platform that can perform differential expression analysis, cluster analysis and PPI network analysis for multi-omics data, but only the PPI network is interactive. GEO2R is a web-based tool for differential expression analysis, which is well integrated in the Gene Expression Omnibus (GEO) database (10). However, the figures of GEO2R are not interactive, and it is limited to analyzing the data in GEO. eVITTA (11) is also a popular web-based platform specially designed for analyzing and visualizing the RNA-seq and microarray data in GEO, which contain differential expression analysis, clustering analysis, and enrichment analysis. iDEP (12) is an easy-to-use web-based tool for analyzing RNA-seq data, and it works by seamlessly implementing dozens of R/Bioconductor packages. piNET (13) is a versatile web platform for downstream analysis and visualization of proteomics data. These existing platforms only provide parts of the whole analysis and exploration process of multi-omics data or single omics data with limited interactive functions. It is difficult to achieve a balance between professionalism and usability. Moreover, there is an urgent need to develop platforms that can combine the advantages of survival analysis in clinical research and omics data visualization in basic research, as more large cohort studies are performed by omics technologies today.

Here, we propose ExpressVis, a biologist-oriented interactive web server aiming to help researchers extract biological and medical insights from expression profile data. As shown in Figure 1, ExpressVis tries to solve the above challenges from the following aspects: (i) It supports the analysis of multiple datasets from microarray, RNA-seq, or mass spectrometry-based proteomics. (ii) It provides a comprehensive set of analysis functions for multi-omics data, including differential expression analysis, clustering analysis, functional enrichment analysis, integration analysis and integration with PPI networks and pathway maps. Additionally, it provides a survival analysis function for omics-based cohort studies. These functions are organized into six modules: DiffExp for differential expression analysis, Cluster-Exp for clustering analysis, SurvivalExp for survival analysis, KeggExp for integration with the KEGG pathway, PPI-Exp for integration with PPI networks, and IntegrationExp for integration analysis for multi-omics data. (iii) It provides rich interactive operations to help researchers explore their data easily.

As shown in Figure 2, these interactive operations not only include basic operations in one interface (e.g. zoom and drag in figures, and filter and sort in tables), but also more sophisticated ones that the existing web-based platforms lack, including (i) interactive operations between different interfaces in a module, such as selecting co-expressed genes in a dendrogram heatmap and displaying these genes in an interactive table, and clicking a gene in a table and visualizing this gene's expression in a box plot; (ii) interactive operations between the modules, such as clicking a pathway in the functional enrichment interface and navigating to KeggExp to visualize this pathway, selecting co-expressed genes in ClusterExp and navigating to PPIExp to visualize the PPI network between these genes. With these interactive functions, even biologists without any programming skills can interact with the maps or figures to conveniently derive more biological insights from expression profile data.

To test ExpressVis and show its rich functions, we performed an independent re-analysis of the multi-omics data from 159 patients with Hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) (14). We were able to recapitulate the original findings and visualize the main results in this paper only using ExpressVis, demonstrating its effectiveness in exploring and visualizing multi-omics data.

MATERIALS AND METHODS

Implementation

The architecture of ExpressVis is based on the client-server website design (Figure 3). The backend server focuses on computation and the front-end client focuses on interactive visualization. The client communicates with the server using RESTful APIs. The backend server is deployed in Docker. The Nginx server receives requests from the client and the requests are sent to the restful APIs. The APIs are based on Django and the Django REST framework. The functions of the APIs in Django utilize some third-party packages, including numpy, pandas, scipy, statsmodels, scikit-learn, and lifelines (15). In the functional enrichment API, *P*-values are calculated with hypergeometric distribution. The differential expression analysis API for microarray data and RNA-seq data, and the survival API in Django, send requests to the corresponding servers in the R environment. The R servers are deployed with Plumber in R, and they perform differential expression analysis and survival analysis. Differential expression analysis utilizes several third-party packages, including DESeq2 (16), limma (17) and microarray annotation packages in Bioconductor. The survival analysis utilizes various third-party packages, including tidyverse, maxstats (18) and survival (19).

The front-end client focuses on helping users explore their data with interactive interfaces. The client was de-
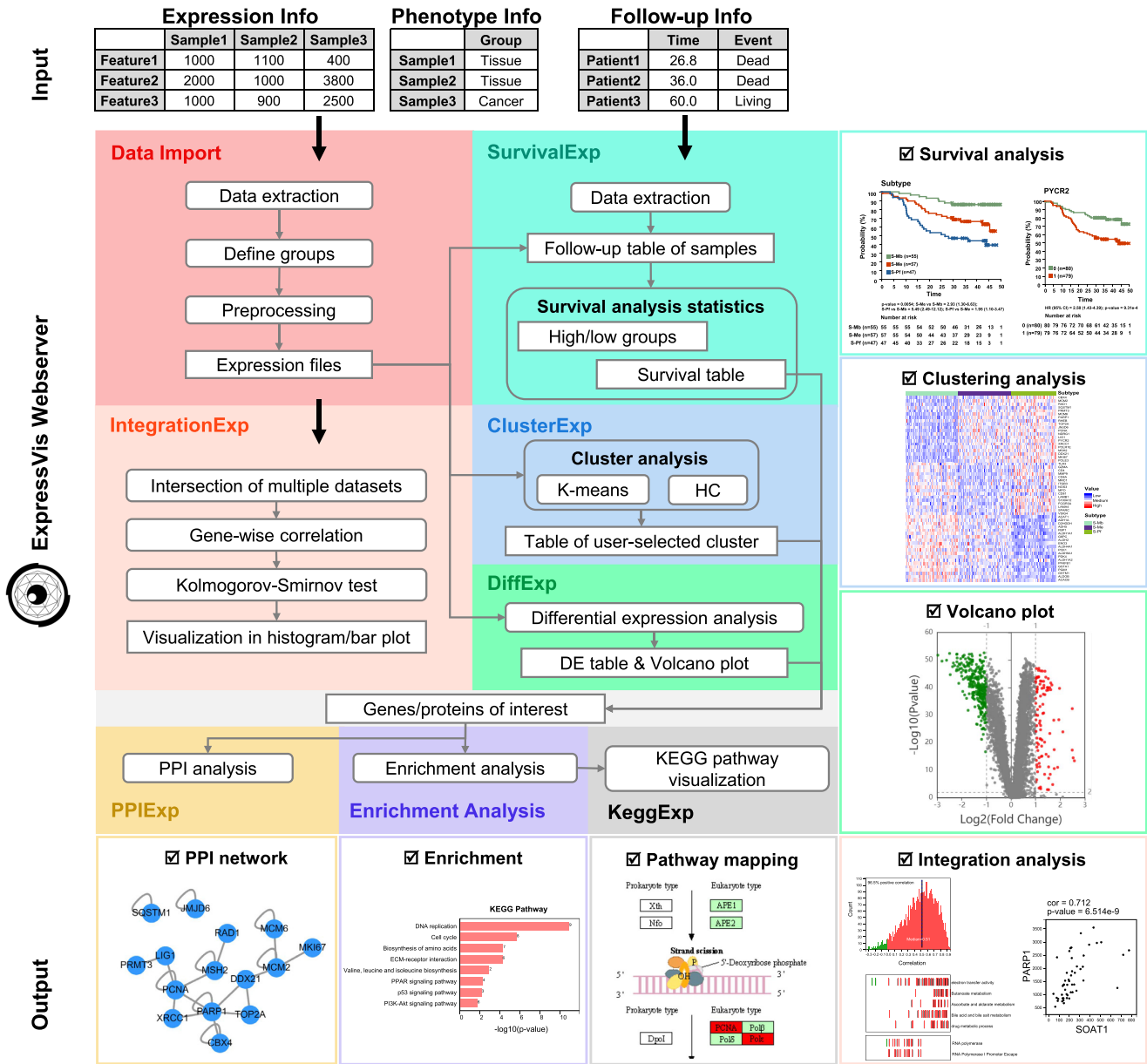
Figure 1. Workflow of ExpressVis. Input files required in ExpressVis are listed on the top. The brief process inside each module and the connection logic between modules are shown in grey lines with an arrow. The right and bottom parts show the representative output of each module in accordance with the color of the module. HC: hierarchical clustering. DE: differential expression.

veloped with Angular 13. To help users explore multiple datasets and gene lists simultaneously and explore analysis results repeatedly, NgRx is used to manage data sets, gene lists, and analysis results. The analysis results from the server and the user's settings are stored in NgRx and users can request them to display and visualize. The results are displayed and visualized with interactive figures and tables. D3.js is used to visualize the expression box plot, bar plot, Kaplan–Meier survival curve (KM curve), forest plot, and heatmap. ECharts is used to visualize a volcano plot. ng-bootstrap is used to display tables. ExpressVis provides basic interactive operations inside the plot and table interfaces. More importantly, ExpressVis provides rich interac-

tive operations among the interfaces. The operations among the interfaces are based on the features of Angular and RxJs.

To make clustering analysis more efficient, WebAssembly is adopted. *K*-means and Hierarchical cluster algorithms are written in C++. The C++ code is compiled to WebAssembly, which can be executed in the browser. The C++ code of hierarchical clustering is modified from fastCluster ([20](20)). Detailed information about the third-party tools used in ExpressVis is listed in Supplementary Tables S1 and S2.

ExpressVis was tested successfully on several browsers, including Chrome, Firefox, Microsoft Edge, and Safari (Supplementary Table S3).
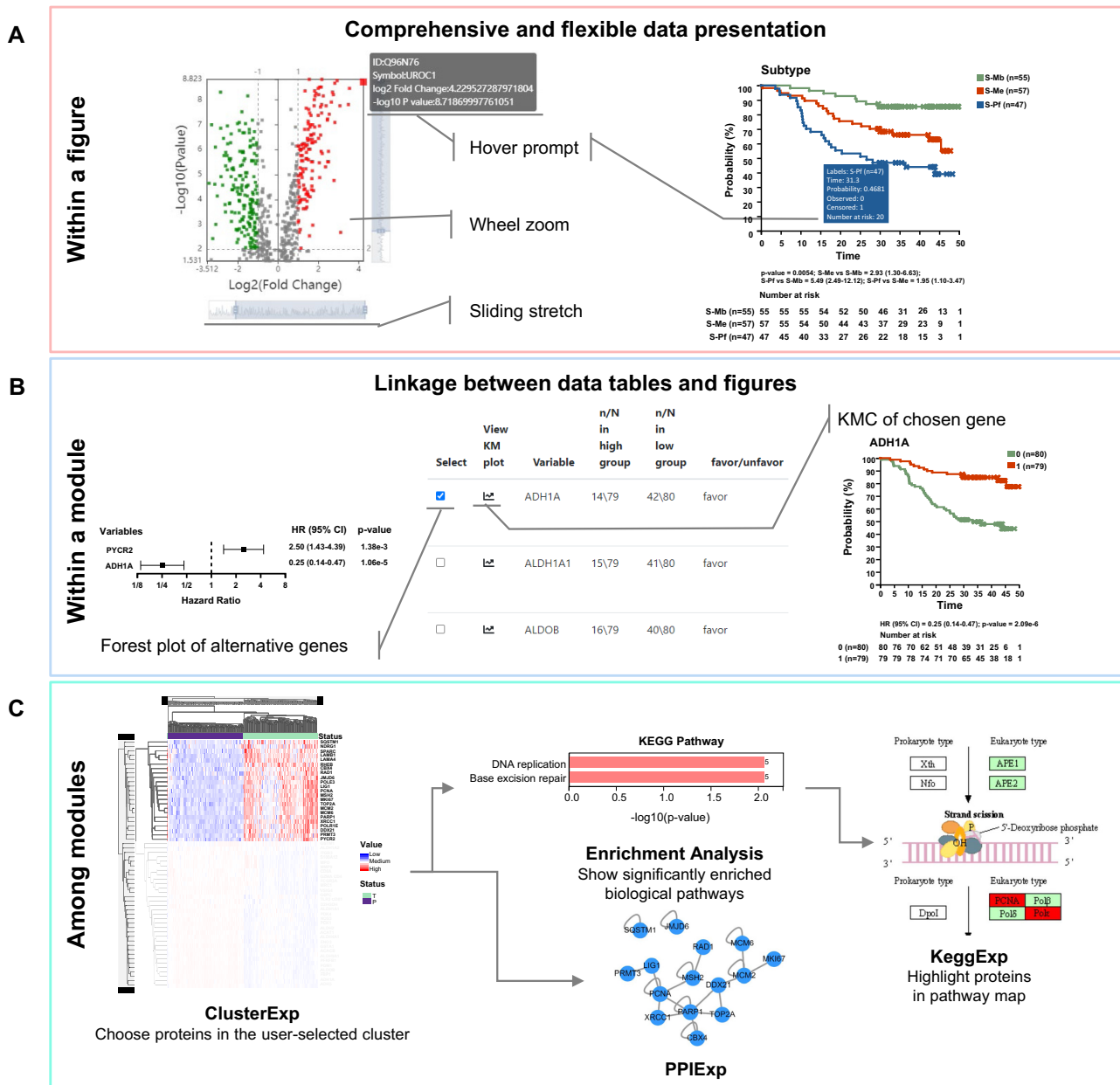
Figure 2. Three levels of interactive operations of ExpressVis. (**A**) The interactive operations within a figure, i.e. hover prompt, wheel zoom, editable text, and the corresponding annotation table for comprehensive and flexible data presentation. (**B**) The linkage between the table and the figure, i.e. select genes/proteins in the table for displaying KM-plot or forest survival plot. (**C**) Interactions among modules: select a cluster of proteins in the expression pattern of interest (ClusterExp) → KEGG pathway enrichment analysis (Enrichment Analysis) → pathway visualization and highlighting (KeggExp).

Deploy ExpressVis with Docker

ExpressVis is deployed with Docker in our Elastic Compute Service in Alibaba Cloud. The deployment process contains only a few commands, making it easy for researchers with limited programming skills. To enable users to use ExpressVis more conveniently, we built a project containing the Docker-compose file and all the necessary source codes to deploy ExpressVis. The project is available on GitHub (https://github.com/OmicsMining/ExpressVis-docker). Users can deploy it easily by following the readme file in GitHub. By deploying ExpressVis using docker lo-

cally or in their server, researchers can explore their data more efficiently.

Data

The database of ExpressVis's backend server mainly stores two types of data, i.e. gene/protein annotation data and Gene Ontology/Pathway annotation data. The gene/protein annotation data are organized by gene/protein ID types. NCBI entrez ID-centered and Refseq Protein ID-centered annotation data are downloaded
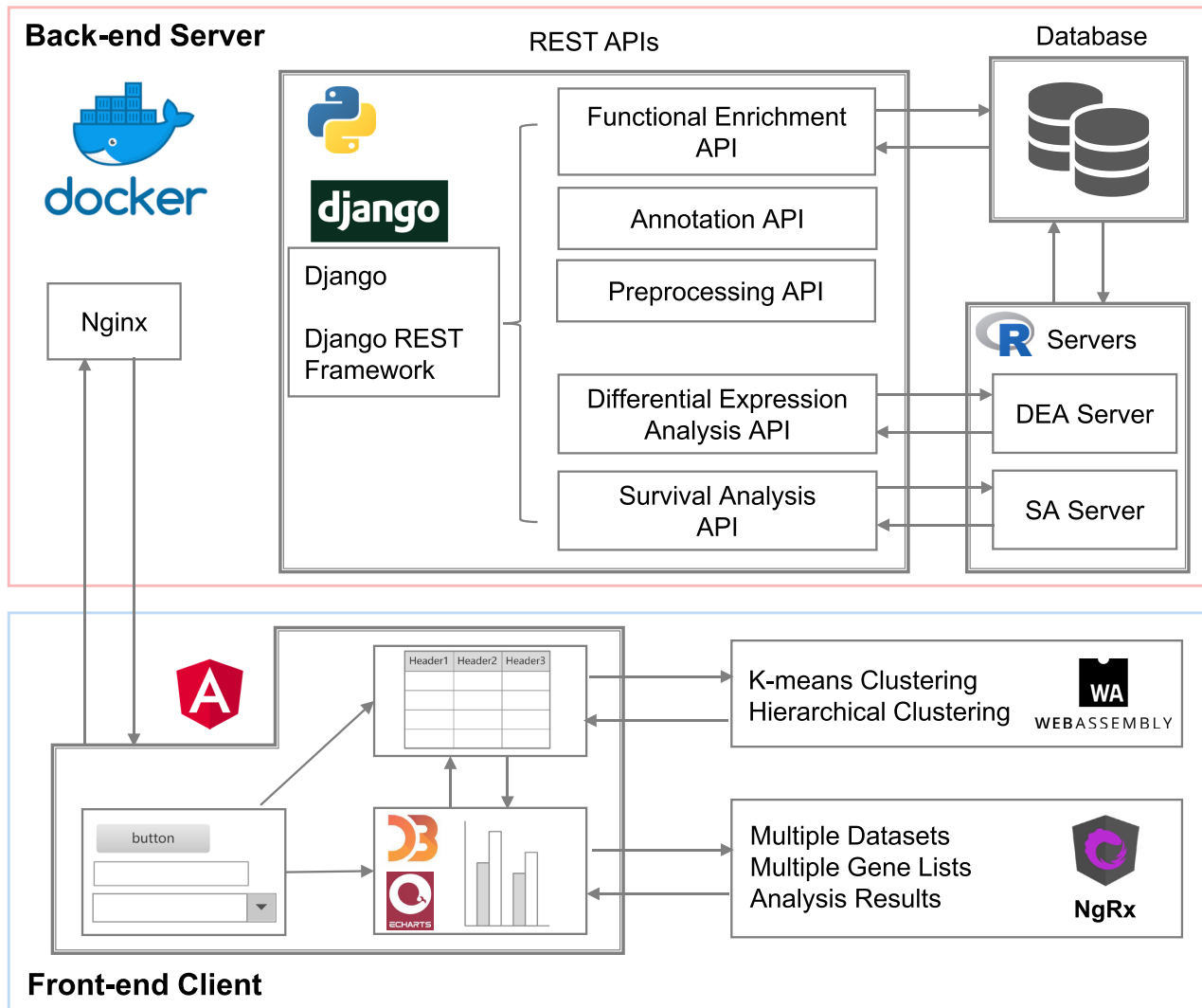
Figure 3. The architecture and implementation of ExpressVis. The solid lines with an arrow represent requests or responses. The boxes in the back-end server represent the servers deployed in Docker. The boxes in the front-end client represent separate parts with different responsibilities. DEA, differential expression analysis; SA, survival analysis.

from NCBI ([21](#)). Ensembl Gene-centered annotation data are downloaded from Ensembl ([22](#)). Uniprot Accession-centered annotation data are downloaded from Uniprot ([23](#)). Microarray annotation data are extracted from microarray annotation packages in Bioconductor. KEGG pathway annotation data are downloaded from the KEGG database ([24](#)). Reactome pathway annotation data are downloaded from the Reactome database ([25](#)). GO annotation data for NCBI entrez ID are downloaded from NCBI. GO annotation data for UniProt Accession are downloaded from UniProt.

Modules

The expression dataset and gene list are two fundamental data types in ExpressVis. An expression dataset contains expression values from multi-omics, phenotype data, and follow-up data. A gene list includes the genes of interest obtained by differential expression analysis, cluster analysis,

survival analysis, or the manually uploaded genes. During the analysis and exploration process, users can add genes of interest into the gene list in ExpressVis flexibly according to their practical needs.

After users import or load one or more datasets, users can use one module independently or use two or more modules interactively. Users can download the whole project and upload the project later to continue the analysis or share it with others. The results of each module are displayed as interactive tables and figures. All of the maps generated by the platform can be downloaded for presentation or publication, and all of the data in the table can be downloaded for later use. The details of these modules are described below.

Data import: compatibility with multi-omics data in formats easily accessible to biologists. ExpressVis supports importing local datasets including microarray data, RNA-seq data, and proteomics data. Importing expression values only requires files that biologists can access easily (Sup-

plementary Table S4 and Supplementary Figure S1). For RNA-seq, the required data are counts and transcripts per kilobase million (TPM)/reads per kilobase million (RPKM)/fragments per kilobase million (FPKM) data. For microarray, the required data are the normalized data from several representative normalization packages, such as Affy (26) and Limma (17). For proteomics, the required data are the quantitative data (without normalization) from typical quantitative software tools, such as MaxQuant (27), Proteome Discoverer, Spectronaut and PANDA (28), and the normalized proteomics data uploaded by users. The interface for preparing the expression dataset provides all the necessary steps for later exploration, including uploading expression data, defining groups, and pre-processing data (such as missing value imputation, data normalization and batch effect correction). The detailed steps are provided in the Supplementary Figure S1. Dataset-related phenotype data and follow-up data are optional. Follow-up data contains patients' follow-up information and is used in SurvivalExp. Batch effect correction algorithms (29,30) were implemented in this module.

DiffExp: identifying differentially-expressed genes/proteins. Differential expression analysis identifies differentially-expressed genes/proteins between two conditions after data normalization and batch effect correction. These genes/proteins can offer biological insights into the processes affected by the conditions of interest. Here, we propose a module named DiffExp to provide complete operations to explore differential expression analysis. Users can repeatedly perform these operations to explore the differential expression analysis results.

As shown in Supplementary Figure S2, DiffExp presents the differential expression analysis results in a volcano plot and an interactive table. The volcano plot is displayed on the left side, and the table is displayed on the right side. In the volcano plot, upregulated genes/proteins are highlighted red, and downregulated ones are highlighted blue. The volcano plot supports a variety of interactive operations to enable users to make the plot printable and explore the plot, including displaying the corresponding data point information (e.g. ID, Gene Symbol, Fold Change and *P*-value) when the mouse hovers over the data point, or changing plot settings, such as height and width. The table displays detailed annotation information of upregulated or downregulated genes/proteins. In the table, users can find genes/proteins of interest through filtering operations. Users can change differential cut-offs, namely the *P*-value or adjusted *P*-value and fold-change, and then the filtered differential genes/proteins are directly shown in the volcano plot and table.

Furthermore, DiffExp provides functional enrichment analysis for the upregulated or downregulated genes/proteins. After the enrichment analysis button is clicked, the module displays the enrichment analysis results, the enriched annotation terms and genes in the annotation terms. The enriched annotation terms are presented with an interactive table and bar plot. Users can interactively select annotation items to visualize in the bar plot. The genes in each annotation term are displayed in the table.

Users can click one term in the table to view the detailed information of the genes in the term.

DiffExp also provides further operations to explore these differentially-expressed genes/proteins. Users can navigate to PPIExp to visualize the PPI network of these genes, navigate to KeggExp to visualize their interested KEGG pathways obtained from the genes' functional enrichment analysis, and add these genes to gene lists for further analysis (Figure 2). In all the tables that display genes/proteins' information, users can click on a gene to view the gene's expression across groups in a boxplot.

ClusterExp: discovering genes with the same expression patterns of interest. Cluster analysis is an essential step in the analysis of omics data because the genes/proteins in the same cluster are likely to have similar biological functions. Here, we propose a module named ClusterExp to help biologists with limited programming skills find co-expressed genes/proteins with similar expression patterns conveniently. This module supports two commonly used clustering methods: hierarchical clustering and *k*-means clustering.

As shown in Supplementary Figure S3, users can click the genes in the root nodes of the dendrogram (for hierarchical clustering) or select co-expressed genes by clicking the heatmap (for *k*-means clustering). Then, they can perform functional enrichment analysis for the selected genes (the interfaces and operations for functional enrichment analysis are similar to those in DiffExp). By repeatedly performing these operations, users can discover genes of interest with similar expression patterns. Additionally, many visualization techniques are applied to make the maps more interactive so as to help researchers better understand high-complexity omics data. Users can drag and zoom in on the map to look for the genes with the expression pattern of interest.

SurvivalExp: survival analysis for omics-based cohort studies. Omics-based technologies are increasingly being used in large-scale cohort studies. Survival-associated genes/proteins have more opportunities to be considered as potential biomarkers and drug-targets, which help researchers gain insights into a disease and guide treatment decisions. The SurvivalExp module was designed to identify the risk and protective factors of concerned genes/proteins via survival analysis.

As shown in Supplementary Figure S4, users need to specify the follow-up information columns such as the survival time, status of the end of the study, and endpoint of interest in status (such as recurrence or death). SurvivalExp can perform statistical analysis on two types of data: (i) for clinical features (e.g. tumor stage and diameter range), it shows the survival differences between groups in the form of a KM curve (KMC); (ii) for genes/proteins of interest, patients are grouped into a high group and a low group according to the gene's expression values by one of the two methods: (a) directly setting the proportion of low groups and (b) screening to find the minimum log-rank *P*-value between the high group and low group based on the maxstat_test method in the R package 'coin'. Survival-related parameters such as the hazard ratio (HR) are calculated,
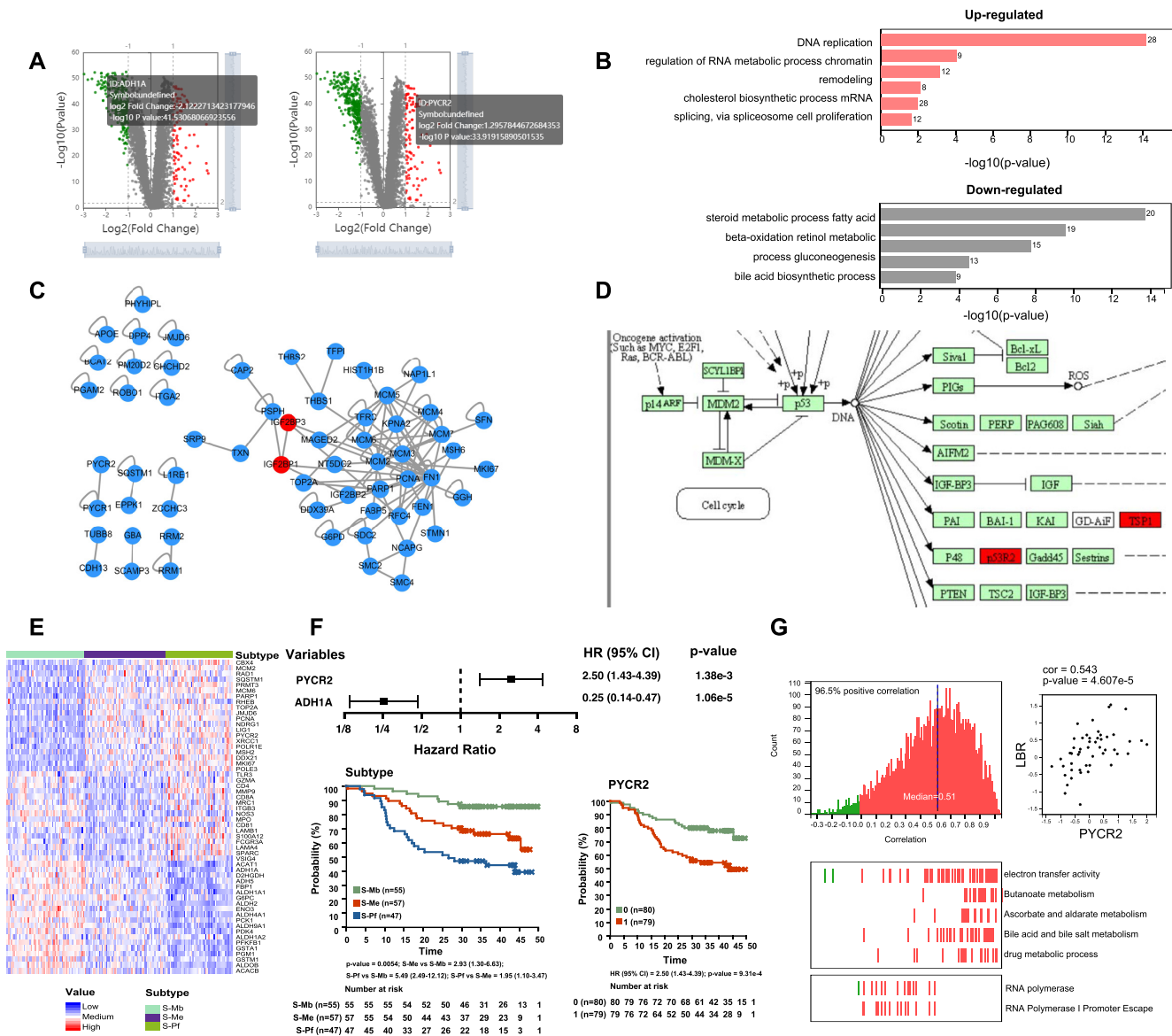
Figure 4. HBV-related HCC proteomics data analysis results. (**A**) Differential expression analysis of tumor (T) versus para-cancer (P) is shown in volcano plots. Up-regulated genes (ratio > 2, *P* < 0.01) are shown in red and down-regulated genes (ratio < 0.5, *P* < 0.01) are shown in green. PYCR2 and ADH1A have the same expression trends in the original paper. (**B**) The enrichment analysis bar plots of T versus P using the gene sets related to 'GO: Biological Process'. The upregulated gene sets are shown in orange and the downregulated gene sets are shown in grey. (**C**) The PPI network map of up-regulated genes of tumor. (**D**) The pathway map depicts up-regulated genes in p53 signaling pathway. (**E**) The heatmap generated in ClusterExp. Each row indicates a protein and each column indicates a tumor sample. (**F**) Survival analysis based on proteomics subtypes or the expressions of PYCR2 and ADH1A in tumor samples (*N* = 159). (**G**) The integration analysis of multi-omics data (proteomics data and transcriptomics data in this case). The gene-wise correlations are plotted in a histogram (up). The genes in the user-selected terms are plotted at the bottom. Red color and green color in the figure indicate positive and negative correlations, respectively. The scatter plot shows the correlation between two genes selected by users.

and the results are displayed in the form of a forest plot and KMC.

Updates on KeggExp and PPIExp. We developed KeggExp (31) and PPIExp (32) in our previous works focusing on integrating expression profile data with KEGG pathways and the PPI network. With the original versions of KeggExp and PPIExp, users can visualize a KEGG pathway and a PPI network, dynamically highlight differentially-expressed genes or specific genes in a pathway map or a PPI network map, and visualize the expressions of genes/proteins of interest with a dendrogram. Herein, we updated KeggExp and PPIExp on the following features: (i) If users import multiple datasets, they can visualize the expression profile of the genes/proteins in the pathway or the PPI network in multiple datasets. (ii) Users can visualize a pathway or a PPI network without importing datasets. They can explore gene lists of interest from other experiments, such as Chip-seq. They can highlight genes in the KEGG pathway maps and construct a PPI network

Table 1. Comparison of ExpressVis with existing web servers. Symbols used for feature evaluations with '√' for present, '-' for absent. The URLs for each tool are given below. PTM means post-translational modification

	ExpressVis	OmicsAnalyst	eVITTA	iDEP	piNET
Data input					
Omics type	Multi-omics	Multi-omics	Transcriptomics (RNA-seq, micro-array)	Transcriptomics (RNA-seq)	Proteomics
Multiple datasets analysis	√	-	√	-	-
Downloadable projects (reusable projects)	√	-	-	-	-
Data pre-processing					
Missing value filtering	√	√	-	√	-
Missing value imputation	√	√	-	√	-
Normalization	√	√	-	-	-
Interactive operations					
Interaction in figures	√	Only in network	√	Only in network	√
Interaction in tables	√	-	-	-	-
Interaction in each module	√	-	-	-	-
Interaction among modules	√	-	-	-	-
Functions					
Differential expression analysis	√	√	√	√	-
Cluster analysis	√	√	√	√	-
Enrichment analysis	√	-	√	√	√
Pathway mapping	√	-	-	√	-
PPI network	√	√	-	√	√
Principal component analysis	-	-	-	√	-
Survival analysis	√	-	-	-	-
PTM analysis	-	-	-	-	√

- ExpressVis: https://omicsmining.ncpsb.org.cn/ExpressVis
- eVITTA: https://tau.cmmt.ubc.ca/eVITTA/
- OmicsAnalyst: https://www.omicsanalyst.ca/
- iDEP: http://ge-lab.org/idep/
- piNET: http://www.pinet-server.org

(Supplementary Figures S5 and S6). (iii) When users navigate between different pathways and PPI networks, the zoom setting of the maps stays the same. This feature helps users explore multiple pathways and PPI networks conveniently.

IntegrationExp: integration analysis for multi-omics data. Two or more omics-based technologies are increasingly being used to profile gene/protein expressions in the same set of samples. Integration analysis of multiple-omics data from the same samples can provide additional biological insights. Here, we present the IntegrationExp module that facilitates integration analysis of the multi-omics data.

In the panel of 'Gene-wise correlation analysis', this module firstly calculates the Spearman's correlation of commonly quantified items (genes or proteins) between two selected datasets (i.e. two types of omics data), which will be displayed as a histogram and a bar plot. In addition, parameters such as the average Spearman's correlation coefficient and the proportion of significantly positively correlated genes are shown to assist in understanding the consistency of the two datasets. Then, the gene set enrichment analysis (GSEA) algorithm (33) is used to identify the annotation terms whose genes are likely to be correlated between two types of omics data (such as transcriptomics data and proteomics data). These enriched terms are shown in an interactive table in descending order based on the normalized enrichment score by default. The genes in the user-selected enriched terms can be plotted under the histogram or the

bar plot interactively. Red and green colors in the figure indicate positive and negative correlations, respectively (Supplementary Figure S7A).

In the panel of 'Correlation analysis for the gene of interest', we provide another integration analysis function to explore the correlations between a gene of interest (the target gene) and all the other genes in different omics data (Supplementary Figure S7B). By setting different correlation cutoffs, users can easily find which gene is highly correlated with the gene of interest in one dataset but not in another, and vice versa. It can help biologists explore the possible transcriptional and translational regulations using multi-omics data and extract biological insights conveniently.

Case study: re-analysis of an HBV-related HCC dataset

To test the capacity and performance of ExpressVis for omics studies of clinical cohorts, we re-analyzed a published HBV-related HCC dataset (14), which contains proteomic and transcriptomic expression data of 159 pairs of tumors (T) and para-cancers (P), along with the corresponding follow-up overall-survival information. The proteomics dataset is from the Supplementary Table S1 of this paper (14) and the transcriptomics dataset is the HCC_UQ_FPKM.tsv file (OED094078) from this paper (14). We performed differential expression analysis between T/P groups (up-regulated: ratio > 2, *P*-value < 0.01, down-regulated: ratio < 0.5, *P*-value < 0.01) by DiffExp based on the proteomics expression data in the original paper

as pre-processed input. In line with the original findings, PYCR2 and ADH1A belonged to the high-regulated and low-regulated tumor groups, respectively, despite the difference in the statistical method (Figure 4A). Enrichment analysis showed higher DNA replication, RNA processing, and proliferation, and lower liver physiological functions, such as the retinol metabolic process, the steroid metabolic process, and the bile acid biosynthetic process in HCC (Figure 4B), which are consistent with the findings of the previous studies (34,35). The interaction network of genes highly expressed in tumors is shown in PPIExp, where IGF2BP1 and IGF2BP3 (ratio > 4) are marked by red color (Figure 4C). We highlighted the genes highly expressed in tumors, i.e. THBS1(TSP1) and RRM2 (p53R2), in the KEGG p53 signaling pathway by KeggExp (Figure 4D). Representative proteins exhibited consistent expression patterns in the heatmap, as in the original paper, when the proteomics subtype served as sample annotation via ClusterExp (Figure 4E). Through the SurvivalExp module, we found that proteomic subtypes are significantly associated with overall survival and the stratification of patients for survival differences, using the median cut-off, was significant for PYCR2 and ADH1A (Figure 4F). Users can perform integration analysis of multi-omics data through the IntegrationExp module as shown in Figure 4G. These results suggest that ExpressVis is convenient enough to analyze the omics datasets from a clinical cohort.

DISCUSSION

In summary, ExpressVis was designed to be a biologist-oriented and efficient analysis and visualization tool for multi-omics data. Table 1 lists all the major features of ExpressVis compared with four other similar web servers. First, ExpressVis supports uploading multiple datasets in different formats, which can be useful to analyze multi-omics data at the same time. Second, ExpressVis provides a comprehensive package of user-friendly interactive operations for interpreting multi-omics big data, which can be classified into three levels (i.e. interactive operations in each figure or table, interactive operations within each module, and interactive operations among modules). The user-friendly interactive visualization functions are one of the most important features in ExpressVis. Third, ExpressVis provides survival analysis functions for multi-omics data with follow-up information, which is easy to use for omics-based cohort studies. Fourth, ExpressVis provides a commonly used integration analysis function for multi-omics data, i.e. performing correlation analysis and identifying the annotation terms whose genes are likely to be correlated between two types of omics data. However, like all similar web servers, ExpressVis also has some limitations. Currently, ExpressVis only supports transcriptomics and proteomics data, meaning it does not support genomics or metabolomics data. We will continue to update ExpressVis to overcome these limitations.

DATA AVAILABILITY

ExpressVis is freely accessible at https://omicsmining.ncpsb.org.cn/ExpressVis and does not require login. Its Docker version is freely available in the GitHub repository (https://github.com/OmicsMining/ExpressVis-docker).

REFERENCES

1. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
2. Aebersold,R. and Mann,M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.
3. Keim,D., Andrienko,G., Fekete,J.-D., Görg,C., Kohlhammer,J. and Melançon,G. (2008) In: Kerren,A., Stasko,J.T., Fekete,J-.D. and North,C. (eds). *Information Visualization: Human-Centered Issues and Perspectives*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 154–175.
4. Mougin,F., Auber,D., Bourqui,R., Diallo,G., Dutour,I., Jouhet,V., Thiessard,F., Thiebaut,R. and Thebault,P. (2018) Visualizing omics and clinical data: which challenges for dealing with their variety?*Methods*, **132**, 3–18.
5. O'Donoghue,S.I., Gavin,A.C., Gehlenborg,N., Goodsell,D.S., Hériché,J.K., Nielsen,C.B., North,C., Olson,A.J., Procter,J.B., Shattuck,D.W. *et al.* (2010) Visualizing biological data-now and in the future. *Nat. Methods*, **7**, S2–S4.
6. Shneiderman,B. (1996) Visual languages, 1996. In: *Proceedings., IEEE Symposium on. IEEE*. pp. 336–343.
7. Yi,J.S., Kang,Y.A., Stasko,J. and Jacko,J. (2007) Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Visual. Comput. Graph.*, **13**, 1224–1231.
8. Chang,C., Xu,K., Guo,C., Wang,J., Yan,Q., Zhang,J., He,F. and Zhu,Y. (2018) PANDA-view: an easy-to-use tool for statistical analysis and visualization of quantitative proteomics data. *Bioinformatics*, **34**, 3594–3596.
9. Zhou,G., Ewald,J. and Xia,J. (2021) OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.*, **49**, W476–W482.
10. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
11. Cheng,X., Yan,J., Liu,Y., Wang,J. and Taubert,S. (2021) eVITTA: a web-based visualization and inference toolbox for transcriptome analysis. *Nucleic Acids Res.*, **49**, W207–W215.

12. Ge,S.X., Son,E.W. and Yao,R. (2018) iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*, **19**, 534.

13. Shamsaei,B., Chojnacki,S., Pilarczyk,M., Najafabadi,M., Niu,W., Chen,C., Ross,K., Matlock,A., Muhlich,J., Chutipongtanate,S. *et al.* (2020) piNET: a versatile web platform for downstream analysis and visualization of proteomics data. *Nucleic Acids Res.*, **48**, W85–W93.

14. Gao,Q., Zhu,H., Dong,L., Shi,W., Chen,R., Song,Z., Huang,C., Li,J., Dong,X., Zhou,Y. *et al.* (2019) Integrated proteogenomic characterization of HBV-Related hepatocellular carcinoma. *Cell*, **179**, 561–577.

15. Davidson-Pilon. (2019) lifelines: survival analysis in python. *J. Open Source Softw.*, **4**, 1317.

16. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

17. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

18. Hothorn,T. and Lausen,B. (2003) On the exact distribution of maximally selected rank statistics. *Comput. Stat. Data Anal.*, **43**, 121–137.

19. Therneau,T.M. and Grambsch,P.M. (2000) The cox model. In: *Modeling survival data: extending the Cox model*. Springer, NY, pp. 39–77.

20. Müllner,D. (2013) fastcluster: fast hierarchical, agglomerative clustering routines for r and python. *J. Stat. Softw.*, **53**, 1–18.

21. Sayers,E.W., Bolton,E.E., Brister,J.R., Canese,K., Chan,J., Comeau,D.C., Connor,R., Funk,K., Kelly,C., Kim,S. *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **50**, D20–D26.

22. Howe,K.L., Achuthan,P., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., Bhai,J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.

23. UniProt,C. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.

24. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.

25. Gillespie,M., Jassal,B., Stephan,R., Milacic,M., Rothfels,K., Senff-Ribeiro,A., Griss,J., Sevilla,C., Matthews,L., Gong,C. *et al.* (2022) The reactome pathway knowledgebase 2022. *Nucleic Acids Res.*, **50**, D687–D692.

26. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy–analysis of Affymetrix genechip data at the probe level. *Bioinformatics (Oxford, England)*, **20**, 307–315.

27. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

28. Chang,C., Li,M., Guo,C., Ding,Y., Xu,K., Han,M., He,F. and Zhu,Y. (2019) PANDA: a comprehensive and flexible tool for quantitative proteomics data analysis. *Bioinformatics*, **35**, 898–900.

29. Johnson,W.E., Li,C. and Rabinovic,A. (2007) Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, **8**, 118–127.

30. Zhang,Y., Parmigiani,G. and Johnson,W.E. (2020) ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform*, **2**, lqaa078.

31. Liu,X., Han,M., Zhao,C., Chang,C., Zhu,Y., Ge,C., Yin,R., Zhan,Y., Li,C., Yu,M. *et al.* (2019) KeggExp: a web server for visual integration of KEGG pathways and expression profile data. *Bioinformatics*, **35**, 1430–1432.

32. Liu,X., Chang,C., Han,M., Yin,R., Zhan,Y., Li,C., Ge,C., Yu,M. and Yang,X. (2019) PPIExp: a web-based platform for integration and visualization of protein-protein interaction data and spatiotemporal proteomics data. *J. Proteome Res.*, **18**, 633–641.

33. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad Sci. U.S.A.*, **102**, 15545–15550.

34. Li,H., Rong,Z., Wang,H., Zhang,N., Pu,C., Zhao,Y., Zheng,X., Lei,C., Liu,Y., Luo,X. *et al.* (2020) Proteomic analysis revealed common, unique and systemic signatures in gender-dependent hepatocarcinogenesis. *Biol. Sex Differ.*, **11**, 46.

35. Jiang,Y., Sun,A., Zhao,Y., Ying,W., Sun,H., Yang,X., Xing,B., Sun,W., Ren,L., Hu,B. *et al.* (2019) Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*, **567**, 257–261.