



Contents lists available at ScienceDirect

## Computational and Structural Biotechnology Journal

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

Software/web server article

## IMOPAC: A web server for interactive multiomics and pharmacological analyses of patient-derived cancer cell lines

Ganxun Li<sup>a,b,c,1</sup>, Dongyi Wan<sup>d,1</sup>, Junnan Liang<sup>a,b,c,1</sup>, Peng Zhu<sup>a,b,c,\*</sup>, Zeyang Ding<sup>a,b,c,\*</sup>, Bixiang Zhang<sup>a,b,c,\*</sup><sup>a</sup> Hepatic Surgery Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China<sup>b</sup> Clinical Medical Research Center of Hepatic Surgery at Hubei Province, Wuhan, China<sup>c</sup> Hubei Key Laboratory of Hepato-Pancreatic-Biliary Diseases, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China<sup>d</sup> Nuclear Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

## ARTICLE INFO

## Keywords:

Human cancer cell lines  
Pharmacogenomics  
Multiomics  
Web server  
Visualization

## ABSTRACT

Large-scale multidimensional cancer genomic and pharmacological profiles have been created by several large consortium projects, including NCI-60, GDSC and DepMap, providing novel opportunities for data mining and further understanding of intrinsic therapeutic response mechanisms. However, it is increasingly challenging for experimental biologists, especially those without a bioinformatic background, to integrate, explore, and analyse these tremendous pharmacogenomics. To address this gap, IMOPAC, an interactive and easy-to-use web-based tool, was introduced to provide rapid visualizations and customizable functionalities on the basis of these three publicly available databases, which may reduce pharmacogenomic profiles from cell lines into readily understandable genetic, epigenetic, transcriptionomic, proteomic, metabolomic, and pharmacological events. The user-friendly query interface together with customized data storage enables users to interactively investigate and visualize multiomics alterations across genes and pathways and to link these alterations with drug responses across cell lines from diverse cancer types. The analyses in our portal include pancancer expression, drug-omics/pathway correlation, cancer subtypes, omics-omics (cis-/trans-regulation) correlation, fusion query analysis, and drug response prediction analysis. The comprehensive multiomics and pharmacogenomic analyses with simple clicking through IMOPAC will significantly benefit cancer precision medicine, contribute to the discoveries of potential biological mechanisms and facilitate pharmacogenomics mining in the identification of clinically actionable biomarkers for both basic researchers and clinical practitioners. IMOPAC is freely available at <http://www.hbpdng.com/IMOPAC>.

## 1. Introduction

The complexity and heterogeneity of tumour genomes bring an enormous challenge when developing novel molecular therapeutics [1]. These strategies greatly depend on our understanding of the genetic events that are causally implicated in malignancy [2]. Observations from clinical practice indicate that molecular alterations in the onco-genome strongly impact the effectiveness of anticancer therapies, and dozens of targeted drugs currently utilized in the clinic or developed in clinical trials have not been linked to specific molecular events [3]. To bridge the gap between precision medicine and cancer heterogeneity, it is crucial to systematically shape a comprehensive pharmacogenomic

landscape of patient-derived cancer models at scale. Recent years have witnessed rapidly accumulated large-scale drug sensitivity screens and corresponding multiomics across pancancer to characterize pharmacogenomics, capture the genetic diversity of cancer and identify drug response biomarkers in multiple public databases, including GDSC [4, 5], DepMap [6,7], and NCI60 [8]; together, they present detailed genetic and pharmacological annotations on more than 2000 cancer cell models. Collectively, such data can be used for biological hypothesis testing, facilitating pharmacogenomics mining in the discovery or repurposing of new anticancer therapeutics and the identification of clinically actionable biomarkers. However, it is still significantly challenging for experimental biologists, especially those without

\* Corresponding authors at: Hepatic Surgery Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China  
E-mail addresses: [zhupeng@tjh.tjmu.edu.cn](mailto:zhupeng@tjh.tjmu.edu.cn) (P. Zhu), [zyding@tjh.tjmu.edu.cn](mailto:zyding@tjh.tjmu.edu.cn) (Z. Ding), [bixiangzhang@hust.edu.cn](mailto:bixiangzhang@hust.edu.cn) (B. Zhang).

<sup>1</sup> These authors contributed equally to this work and share first authorship.

<https://doi.org/10.1016/j.csbj.2023.07.023>

Received 20 June 2022; Received in revised form 15 July 2023; Accepted 19 July 2023

Available online 22 July 2023

2001-0370/© 2023 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

bioinformatic experience, to integrate, explore, analyse, and visualize these large-scale pharmacogenomic profiles. Therefore, we developed the IMOPAC, an interactive and easy-to-use webserver for comprehensively characterizing the landscape of pharmacogenomic interactions and easily visualizing the pharmacogenomic profiles in preclinical cancer models.

There are numerous and diverse methods for analysing pharmacogenomics. Clustering analysis, based on multiomics, can identify which cell line models most closely resemble specific cancer subtypes and provide a better understanding of biological mechanisms by directly linking in vitro models to patient tumour samples with curated clinical data available. Liu et al., for instance, divided SNU398, HepG2 and SNU475 cells into three distinct DNA methylation HCC subtypes by unsupervised clustering analysis, which can act as a preclinical model representing different HCC molecular features to investigate the response of HCC cells to DNMTi treatment [9]. Pancancer expression analysis is also widely used to facilitate a comprehensive understanding of a given gene's molecular pattern across multiple different cancer types. Chai et al. revealed that *ACE2* was overexpressed in lung adenocarcinoma by pancancer expression analysis [10]. It is generally recognized that multiomics alterations, including somatic mutation, copy number variation, DNA methylation, transcriptomics, and proteomics, remarkably impact the clinical drug response to treatment and, in numerous situations, are potential biomarkers for sensitivity to chemotherapy [11,12]. Correlation analysis between drug sensitivity and multiomics alterations or pathway activity is applied to facilitate pharmacogenomics mining in the discovery of the underlying biological mechanism of drug resistance. By combining large-scale genomic dependency and pharmaceutical screening datasets within preclinical cell line models, Zhou et al. revealed that the combination of ERBB inhibitors could augment the efficacy of CDK4/6 inhibitor monotherapy in several squamous cell lines both in vitro and in vivo [12]. Cis-regulation analysis is used to investigate the correlation among the multiomics levels of a specific gene at the same time. In recent studies, several DNA methylation-driven genes were identified based on the correlation of gene expression and promoter DNA methylation [13]. Transregulation analysis is applied to estimate the association among various genes from the same or different omics. Because genes with similar expression features have a tendency to possess related functions, it is frequently desirable to detect genes with expression similarities to a known gene, acting as a known tumour chemotherapy target [14]. Gene fusions frequently result from rearrangements in cancer genomes. In various instances, gene fusions play an important role in oncogenesis [15]. A better understanding of the gene fusion detected in cell lines may improve diagnosis and further benefit translational research into targeted therapies necessary to treat these malignancies [16]. In addition, prediction of clinical drug response for cancer drugs is critical for personalized treatment [17].

Currently, CellMinerCDB [18], PharmacoDB 2.0 [19] and RNAactDrug [20] have provided many useful visualization and analysis tools for pharmacogenomic analysis in patient-derived cancer cell lines. CellMinerCDB aims to integrate genomic alterations, molecular signatures and specific drug responses of cancer cell lines. PharmacoDB 2.0 has focused on integrating across a wider range of pharmacogenomics studies and includes major pancancer screening initiatives as well as smaller, tissue-specific studies. RNAactDrug is a comprehensive database of RNAs, including lncRNAs, miRNAs and mRNAs, associated with drug sensitivity from multiomics data, which allows the user to explore drug sensitivity and RNA molecule associations directly. While these databases are exceptionally valuable and widely used, many additional functions are often requested by experimental biologists but are not adequately addressed by existing tools. For example, correlation analyses between customized pathway activity (by user or originating from MSigDB [21]) and drug responses are commonly performed, but this function is not available in CellMinerCDB. In addition, while PharmacoDB 2.0 provides tissue-specific pharmacogenomic analysis, it does not

integrate any gene essentiality data and does not allow to explore any molecular features associated with drug response. Furthermore, correlation analyses between drug response and other omics, including proteomics, metabolomics, and genetic dependency data, the visualization of tissue-specific gene–drug associations, and multivariate analysis of drug response with respect to multiple-omic features simultaneously are not available in RNAactDrug. None of the existing databases allow specific analyses based on cis-regulation (the correlation among the multiomics level of a specific gene) and trans-regulation (the association among various user-input genes from the same or different omics). The effect of DNA methylation in promoter versus gene body CGIs on its corresponding gene expression remains contradictory [22]. The pharmacogenomic analyses based on gene body or promoter DNA methylation are not available in previous databases. Moreover, none of the databases provide a fusion query, molecular classification across pancancer or within specific cancer types, or pancancer expression analysis in patient-derived cancer cell lines. Based on the abovementioned needs that are not adequately addressed, we developed IMOPAC, a web server to deliver fast and customizable functionalities to complement with the existing tools.

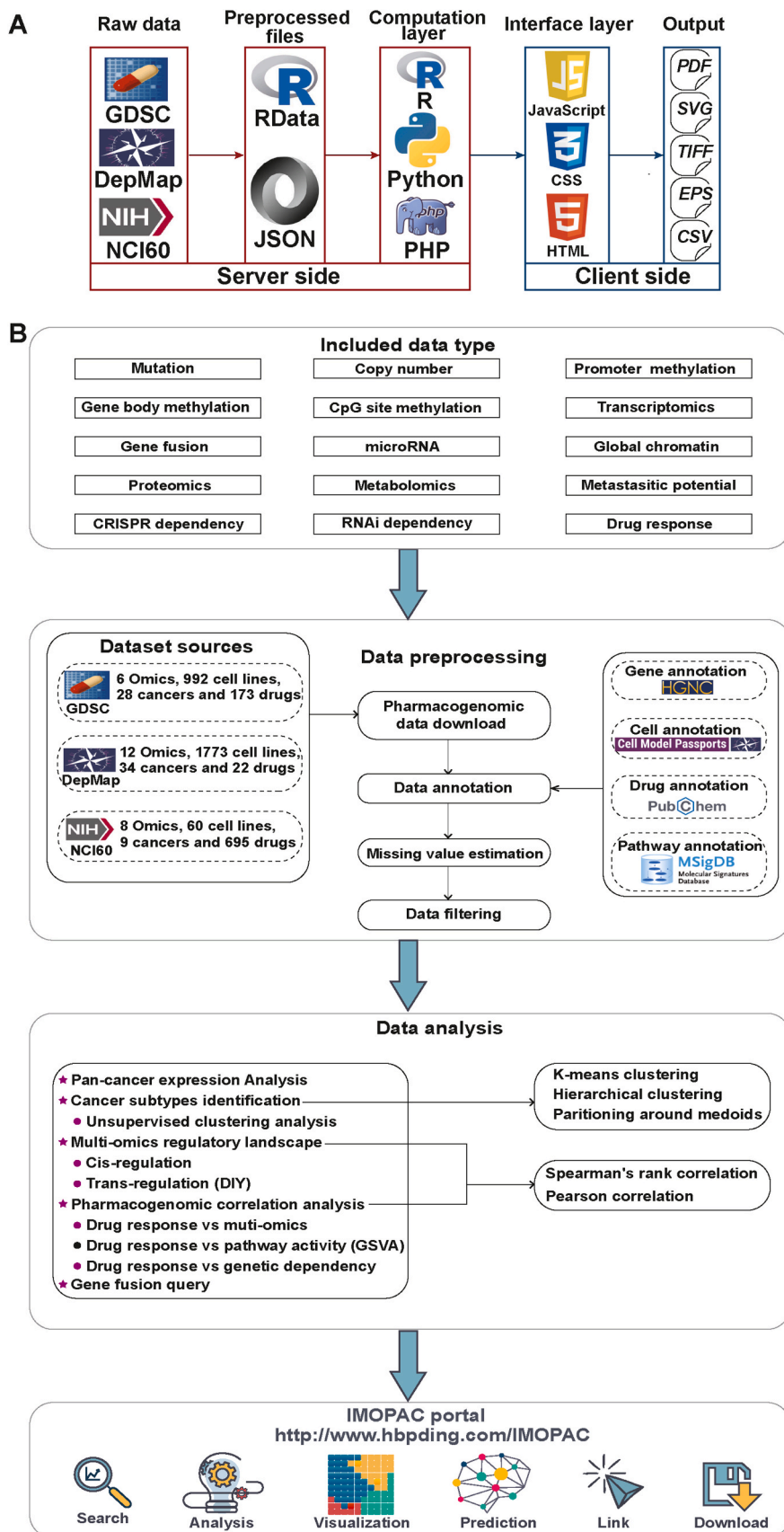
## 2. Materials and methods

### 2.1. Web interface implementation

The IMOPAC website is free for all users, and no login is required to access all functions. We run the IMOPAC webserver on Apache (v2.4.38, <https://apache.org/>) with the PHP module as the httpd service on Debian 10. Figures shown on the website are in the svg format generated by SVG.js (v2.7, <https://svgjs.dev/docs/2.7/>). Backend analyses and figure drawing were performed in R (v4.1.2, <https://www.r-project.org/>). The IMOPAC web server is compatible with browsers (on Windows, MacOS or Linux) that support HTML5 and CSS3. All pages were written in HTML5, CSS3, and original JavaScript (Fig. 1A). All IMOPAC source codes (back-end, front-end and data processing) can be downloaded from the GitHub repository: <https://github.com/wolfgangsk07/IMOPAC>.

### 2.2. Data collection and preprocessing

IMOPAC is based on multiomics and drug response screens originating from three publicly available databases (6 omics, 992 cell lines from 28 cancer types and 173 drugs are available in GDSC; 12 omics, 1773 cell lines from 34 cancer types and 22 drugs are available in DepMap database; 8 omics, 60 cell lines from 9 cancer types and 695 drugs are available in NCI60) (Table 1). A flowchart of the IMOPAC pipeline is shown in Fig. 1B. The input gene-related CpG sites were annotated by the *IlluminaHumanMethylation450kanno.ilmn12.hg19* R package [13]. The DNA methylation profiles of the gene promoter and body were generated by calculating the average DNA methylation value of CpG sites in the promoter and body region of the specific gene, respectively. Detailed clinical and biological features of all cell models, including age, sex, ploidy, microsatellite status, mutational burden, culture type, and culture medium, were retrieved from Cell Model Passports [23] and DepMap [6]. Whole genes and drugs in this database were annotated by the HGNC [24] and pubchem [25] databases, respectively. We excluded drugs whose sensitivity values were 'NA' in greater than 15 % of the cell lines. The gene fusion profile was estimated based on RNA-seq data from DepMap Public 22Q1. Data were filtered using the following: 1. removing fusion involving mitochondrial chromosomes or HLA genes; 2. removing common false-positive fusions (red herring annotations as described in the STAR-Fusion docs); 3. recurrent fusions observed in DepMap across cell lines (in 10 % or more of the samples); 4. removing fusions where SpliceType = 'INCL\_NON\_REF\_SPLICE' and LargeAnchorSupport = 'NO\_LDAS' and FFPM < 0.1.



**Fig. 1. Overview of IMOPAC.** A. IMOPAC software architecture; B. IMOPAC integrates cancer cell line information from multiple sources and provides ready-made, user-friendly analysis tools as well as data download features for further analyses.

**Table 1**  
The available omics in three public datasets.

The available omics in three public datasets			
	GDSC	DepMap	NCI60
Mutation	Yes	Yes	Yes
CNV	Yes	Yes	Yes
mRNA	Yes	Yes	Yes
DNA methylation (Body)	Yes	No	Yes
DNA methylation (Promoter)	Yes	Yes	Yes
DNA methylation (CpG site)	Yes	No	Yes
Fusion	No	Yes	No
microRNA	No	Yes	Yes
Protein	No	Yes	Yes
Metabolomics	No	Yes	No
Global Chromatin	No	Yes	No
Metastatic Potential	No	Yes	No
RNAi dependency	No	Yes	No
CRISPR dependency	No	Yes	No

### 2.3. Visualization and statistical analysis

Every figure provides not only a visualization of the results of pharmacogenomic analyses but also several statistical analyses. The numbers on the right side of the correlation plot present the P/FDR values and Pearson/Spearman correlation coefficients estimated by comparing the differences between the reference (located in the bottom line) and all other variables for users. When multiple hypothesis testing is selected for statistical analysis, Benjamini–Hochberg adjusted *p*-values will be provided by our web server.

#### 2.3.1. Pancancer expression analysis

A histogram is used to assess the alteration frequency for omics with discrete settings (e.g., CNV or somatic mutation), and a boxplot is used to investigate the expression abundance for those with continuous settings (e.g., mRNA and so on) of a specific input gene across multiple cancer types in the “Cancer Types Summary” module of IMOPAC.

#### 2.3.2. Drug-omics correlation analysis

To investigate the association between multiomics molecules and drug responses hidden in large-scale pharmacogenomics, two different statistical analyses were performed to identify this correlation across user-defined cell lines according to data types.

First, for omics with discrete settings (e.g., somatic mutation), we first divided cell models into two subgroups according to the mutation status of a given gene and performed a user-defined statistical analysis, including Wilcoxon's sign rank test and Student's *t*-test, on the drug response data of these two subgroups. A significant difference in pharmacological profiles between these two subgroups suggested that the mutation status of this specific gene has an impact on drug sensitivity [26].

Second, for omics with continuous settings (e.g., mRNA and so on), a user-defined correlation analysis, including Pearson and Spearman methods, was performed to evaluate the correlation between drug response and molecular expression values of multiomics molecules.

The user-defined parameters, including the type of omics, input gene, drugs, cell lines, threshold values of correlation coefficient, statistical method of correlation analysis, *P*-values and adjusted *P*-values were used to obtain significant results. All correlations between molecular alterations of the user-defined gene and the response of user-defined drugs were visualized in the “Drug-Omics” module of IMOPAC.

#### 2.3.3. Drug-pathway correlation analysis

A user-defined correlation analysis was performed to evaluate the correlation between drug response and the activity of the specific pathway, which was calculated by gene set variation analysis [27] based on a transcriptomic geneset originating from user customization or MSigDB [21].

The user-defined parameters, including transcriptomic geneset, drugs, cell lines, threshold values of correlation coefficient, statistical method of correlation analysis, *P*-values and adjusted *P*-values were used to obtain the specific result. All correlations between the activity of a specific pathway and the response of user-defined drugs were visualized in the “Drug-Pathway” module of IMOPAC.

#### 2.3.4. Cancer subtype analysis

Cancer subtype analysis could provide an unsupervised clustering analysis for identifying cancer subtypes based on a geneset from user-defined multiomics. IMOPAC provides three feature selection methods for unsupervised clustering analysis, with distinct metric and linkage parameters available, to identify the key features in a multiomics dataset. The “Cancer Subtype” module presents a heatmap based on the results of cancer subtype analysis, in which the samples are separated into 2–4 subgroups according to the user's choice.

#### 2.3.5. Cis-regulation analysis

Cis-regulation analysis was performed to investigate the correlation among the multiomics levels of a specific user-input gene. The mRNA expression of this input gene is generally chosen as a reference.

If the other variable is also numeric (for example, the expression level of microRNA), IMOPAC will estimate the correlation between the reference and the microRNA data through the Pearson or Spearman method according to the user demands. If the other variable is categorical, IMOPAC will perform a Wilcoxon's sign rank test or Student's *t*-test to check whether there is a significant difference in expression between these two groups. The user-defined parameters, including input genes, cell lines, threshold values of correlation coefficients, statistical methods of correlation analysis, *P*-values and adjusted *P*-values were used to obtain the results, which were visualized in the “Omics-Omics (cis-regulation)” module of IMOPAC.

#### 2.3.6. Trans-regulation analysis

Trans-regulation analysis is used to estimate the association among various user-input genes from the same or different omics.

When a numeric variable is chosen as a reference, IMOPAC will go over all other variables in the figure and perform an appropriate test. If the other variable is also numeric (for example, the expression level of microRNA), IMOPAC will estimate the correlation between the reference and the microRNA data. If the other variable is categorical, IMOPAC will perform a Wilcoxon's sign rank test or Student's *t*-test to check whether there is a significant difference in expression between these two groups. In contrast, the chi-square test will be utilized when the reference and the other data are all categorical, and the Wilcoxon's sign rank or Student's *t*-test will be adopted when the reference is categorical and the other variables are numeric. The user-defined parameters, including input genes, type of omics, cell lines, threshold values of correlation coefficient, statistical method of correlation analysis, *P*-values and adjusted *P*-values, were used to obtain the results, which were visualized in the “Omics-Omics (transregulation)” module of IMOPAC.

#### 2.3.7. Performing fusion queries

A Circos plot was used to visualize the comprehensive gene fusion landscape of the user-defined cell line in the “Fusions” module of IMOPAC by the R package *circize*[28].

### 2.4. Predicting drug response by machine learning algorithms

Ten algorithms with 5-fold stratified cross-validation were provided in our portal based on the GDSC dataset, including random forest performed by *randomForest* R package, neural networks performed by *nnet* R package, adaboost performed by *adabag* R package, gradient boosting machine performed by *caret* R package, support vector machines performed by *e1071* R package, decision tree performed by *rpart* R package, k-nearest neighbor performed by *kknn* R package, ridge regression

performed by *glmnet* R package, lasso regression performed by *glmnet* R package, logistic regression performed by *glm* R base function, which were used to construct model and predict the response of user-defined drug with default hyperparameters in the “Machine Learning in Predicting Drug Response” module of IMOPAC. The KNN method from *impute* R package was used to impute the deletion value of the multiomics profiles. The predictive model used normalized multiomics signature by the scale method as the input and the sensitive or resistant label of cell lines as dependent variable. All R packages were downloaded from <https://cran.r-project.org/web/packages/>. For a specific model, the average AUC ( Area Under Curve ) of selected drug was calculated to investigate the predictive performance of user-defined multiomics signature. In addition, the available coefficients of the specific model will be presented in the pop-up windows.

### 2.5. More features of the web server

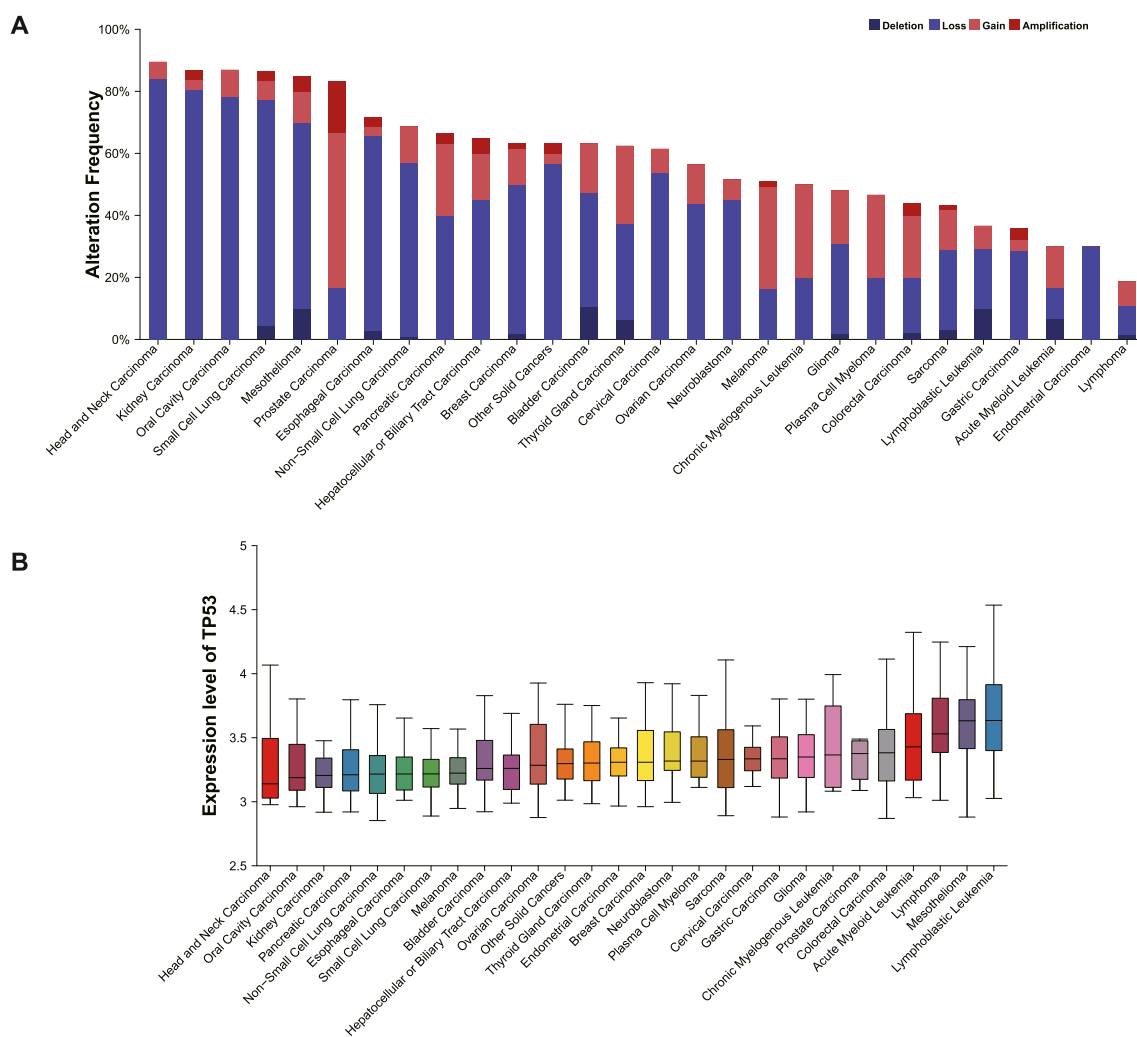
1. Automatic prompt and completion function are applied when inputting genes;
2. More detailed information on the drug, cell line, and CpG site can be viewed in the pop-up windows when moving mouse over any cell line, drug and CpG site's text or any bar in the histograms and boxplots;

3. Web links, located in the pop-up windows, are generated to obtain more detailed annotations of cell lines and drugs from external databases and resources;
4. All figures are presented in print size, which is marked on the top and left of the plot area;
5. When the figure is large and complex, the user can reduce the size of the figure by zooming on a specific part. To do this, the user has to click “+” at the top of the plot area. The user can also zoom out again by clicking the “-” button and reset the size by clicking the double circle button. Zooming does not affect the downloading of the figure;
6. The outputs from all analyses, including tables (csv) and high-resolution figures (pdf, tiff, svg, and eps), can be easily downloaded from IMOPAC.

## 3. Results

### 3.1. Home

IMOPAC provides a graphical interface that briefly introduces information for all three datasets used in this project and is a web server that aims to provide interactive multiomics and pharmacological analyses between multiomics molecules, cells and drugs in human cancers. As shown in Fig. 1B, seven modules based on three datasets were provided: (i) pancancer expression analysis, (ii) drug-omics correlation



**Fig. 2.** Examples of pancancer expression analysis in the “Cancer Types Summary” module. By clicking the “START” button, this module will present a histogram (A) for input omics with discrete settings (the CNV status of *CTNNB1*), which shows the frequency of the variable in descending order across cancers, as well as a boxplot (B) for input omics with continuous features (the mRNA level of *TP53*), which shows the expression abundance in ascending order across cancers.

analysis, (iii) drug-pathway correlation analysis, (iv) cancer subtype analysis, (v) omics-omics (cis-regulation) correlation analysis, (vi) omics-omics (transregulation) correlation analysis, and (vii) fusion analysis.

### 3.2. Pancancer expression analysis

Pancancer expression analysis is generally performed to investigate the molecular pattern of a specific gene across multiple different cancer types. Accordingly, the “Cancer Types Summary” module will present a histogram for input omics with discrete settings, which shows the mutation frequency of the input gene in descending order across cancers, or a boxplot for input omics with continuous features, which shows the expression abundance in ascending order across cancers. The example shows that *CTNGB1* CNV alteration frequencies are the highest in head and neck carcinoma, kidney carcinoma, and oral cavity carcinoma (Fig. 2A), and the mRNA level of *TP53* is the highest in lymphoma, mesothelioma, and lymphoblastic leukaemia in the GDSC database (Fig. 2B).

### 3.3. Drug-omics correlation analysis

Correlation analysis between drug sensitivity and multiomics alterations or pathway activity is applied to facilitate pharmacogenomics mining in the discovery of the underlying biological mechanism of drug resistance or repurposing of new anticancer therapeutics and the identification of clinically actionable biomarkers. Therefore, in the “Drug-Omics” module, IMOPAC allows users to investigate the correlation between the omics level of user-input genes and the response of all selected drugs across user-defined cell lines. This module will present a correlation plot based on the user’s custom parameters, in which the samples are ordered ascendingly by the omics level of the input and the statistical results are shown on the right-hand side. For example, users can examine the correlation coefficients between the mRNA level of *TP53* and 17 drugs targeting the PI3K/MTOR signalling pathway (Fig. 3A) and the association between the mutation status of *TP53* and 15 drugs targeting the DNA replication signalling pathway across all

selected cell lines based on the GDSC dataset (Fig. 3B).

### 3.4. Drug-pathway correlation analysis

The drug-pathway module is designed to explore the association between the activity of a specific pathway and the response of all selected drugs across user-defined cell lines. This module will present a correlation plot based on users’ custom parameters, in which the samples are ordered ascendingly by the activity of the specific pathway, and the statistical results are shown on the right-hand side. For example, users can examine the correlation coefficients between the activity of HALLMARK\_ANGIOGENESIS and 7 drugs targeting EGFR signalling (Figs. 4A and 5) and the correlation coefficients between the activity of hypoxia signalling calculated by the user’s customized geneset and 17 drugs targeting PI3K/MTOR signalling (Fig. 4B) across all selected cell lines based on the GDSC dataset.

### 3.5. Cancer subtype analysis

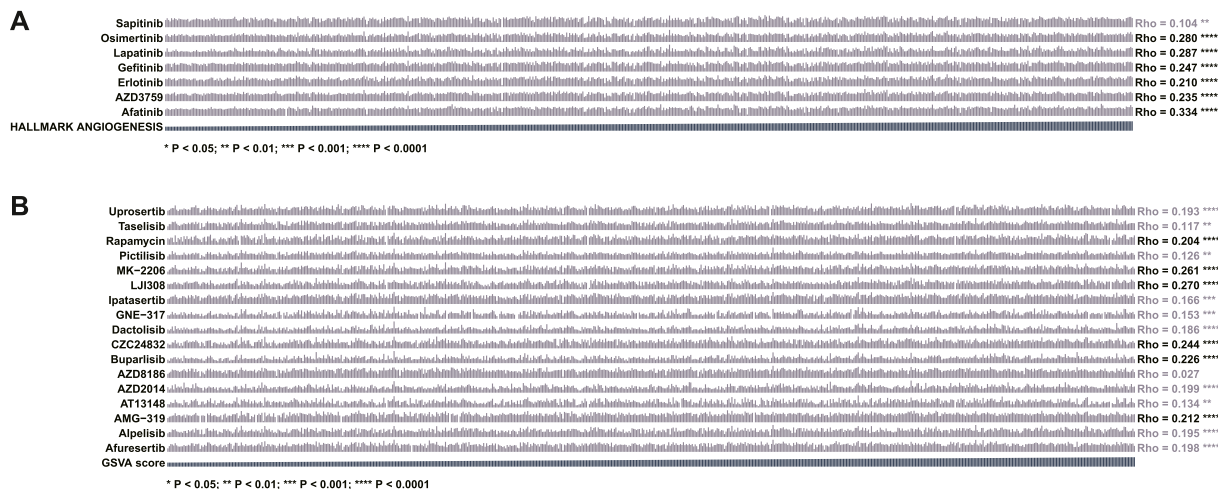
Cancer subtype analysis, based on multiomics, can identify which cell line models most closely resemble specific cancer subtypes and provide a better understanding of biological mechanisms by directly linking in vitro models with patient tumour samples with curated clinical data. Consequently, the “Cancer subtype” module is developed to provide an unsupervised clustering analysis for identifying cancer subtypes from multiomics raw data to result visualization based on a user-defined geneset. In the heatmap, red colour indicates a high expression pattern, whereas blue colour indicates a low expression pattern. Meanwhile, the clustering result of cell models will be presented in tabular form when there is not enough space to list the names of cell lines under the heatmap. For example, users can obtain three specific subgroups by selecting specific metric (euclidean) and linkage (ward.D) parameter across all selected cell lines based on the DepMap dataset.

### 3.6. Omics-Omics (cis-regulation) correlation analysis

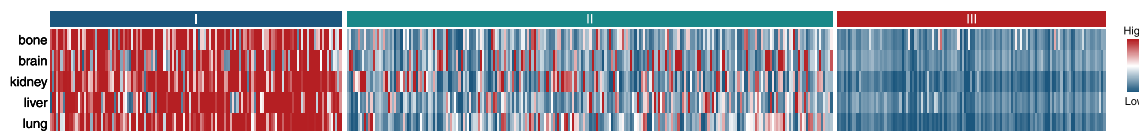
In the “Omics-Omics (cis-regulation)” module, cis-regulation



**Fig. 3.** Examples of drug-omics correlation analysis in the “Drug-Omics” module. By clicking the “START” button, this module will present a correlation plot based on users’ custom parameters, in which the samples are ordered ascendingly by the omics level of the input; white colour is used to present NA values, and the statistical results (for input omics with continuous settings (A), black colour represents that the absolute correlation coefficients are more than the user-defined cut-off. When the data type of input omics is discrete (B), black colour represents that *P* or FDR values are less than 0.05) are shown on the right-hand side. The IC50 values of all user-defined drugs illustrated in the figure were normalized by deviation standardization.



**Fig. 4.** Examples of drug-pathway correlation analysis in the “Drug-Pathway” module. By clicking the “START” button, this module will present a correlation plot based on users’ custom parameters, in which the samples are ordered ascendingly by the activity of the selected pathway. The activity of the specific pathway was calculated by gene set variation analysis based on a transcriptomic gene set originating from user customization (A) or MSigDB (B). White colour is used to present NA values, and the statistical results (black represents that the absolute correlation coefficients are more than the user-defined cut-off) are shown on the right-hand side. The IC50 values of all user-defined drugs illustrated in the figure were normalized by deviation standardization.



**Fig. 5.** Examples of cancer subtype analysis in the “Cancer subtype” module. By clicking the “START” button, this module will present a heatmap based on the unsupervised clustering analysis, with distinct metric and linkage parameters available. In these results, red colour indicates a high expression pattern, whereas blue colour indicates a low expression pattern.

analysis is performed to investigate the correlation among the multi-omics level of a specific user-input gene at the same time. For example, the correlation among the mRNA, CNV, mutation status, DNA methylation of CpG sites located in *TP53*-related regions, TMB, ploidy, and MSI status across all selected cell lines based on the GDSC dataset is visualized in Fig. 6.

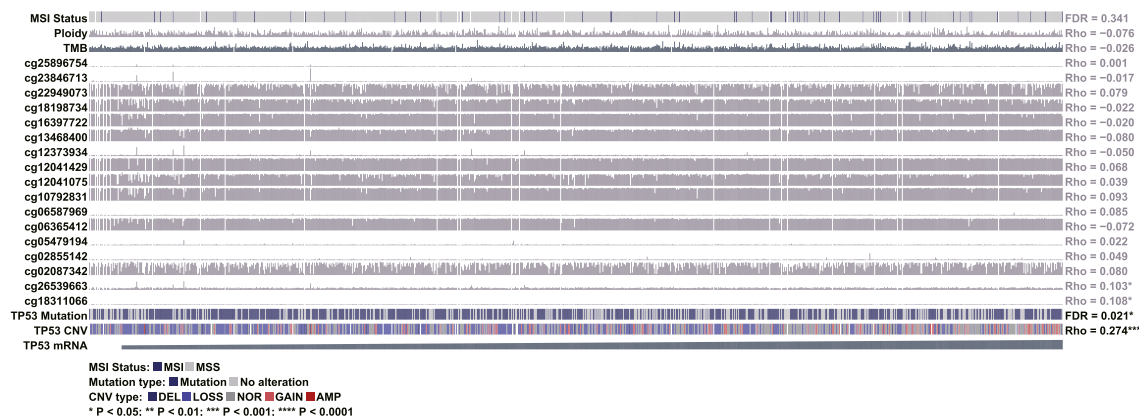
3.7. Omics-Omics (transregulation) correlation analysis

In the “Omics-Omics (transregulation)” module, trans-regulation

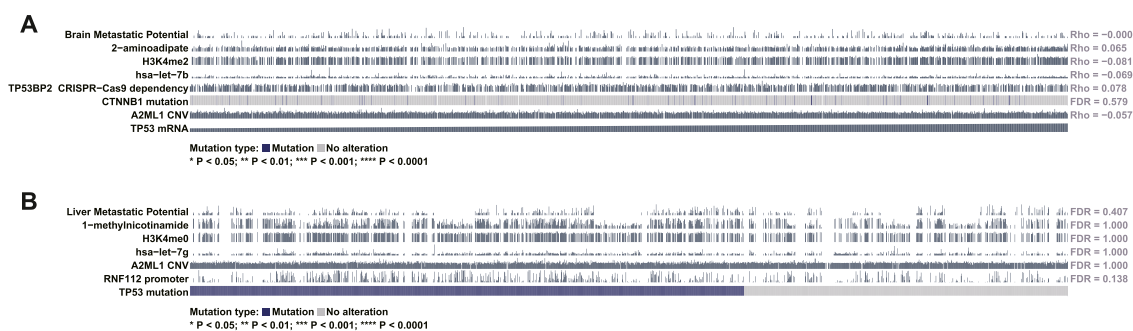
analysis is used to estimate the association among various user-input genes from the same or different omics. For example, a numeric variable (the mRNA level of *TP53*) was chosen as a reference, and the correlation is visualized in Fig. 7A. Meanwhile, the reference can also be categorical (the mutation status of *TP53*), and the association is visualized in Fig. 7B.

3.8. Fusion analysis

Gene fusions frequently result from rearrangements in cancer



**Fig. 6.** Examples of omics-omics (cis-regulation) correlation analysis in the “Omics-Omics (cis-regulation)” module. By clicking the “START” button, this module will present a cis-regulatory correlation plot based on the input gene, in which samples are ordered ascendingly by the mRNA level of this gene, white colour is used to present NA values, and the statistical results (for input omics with continuous settings, black colour represents that the absolute correlation coefficients are more than the user-defined cut-off; when the data type of input omics is discrete, black colour represents that *P* or *FDR* values are less than 0.05) are shown on the right-hand side. Except for DNA methylation profiles, the omics levels of variables with continuous settings and the values of ploidy and log<sub>2</sub>(TMB) illustrated in the figure are normalized by deviation standardization.

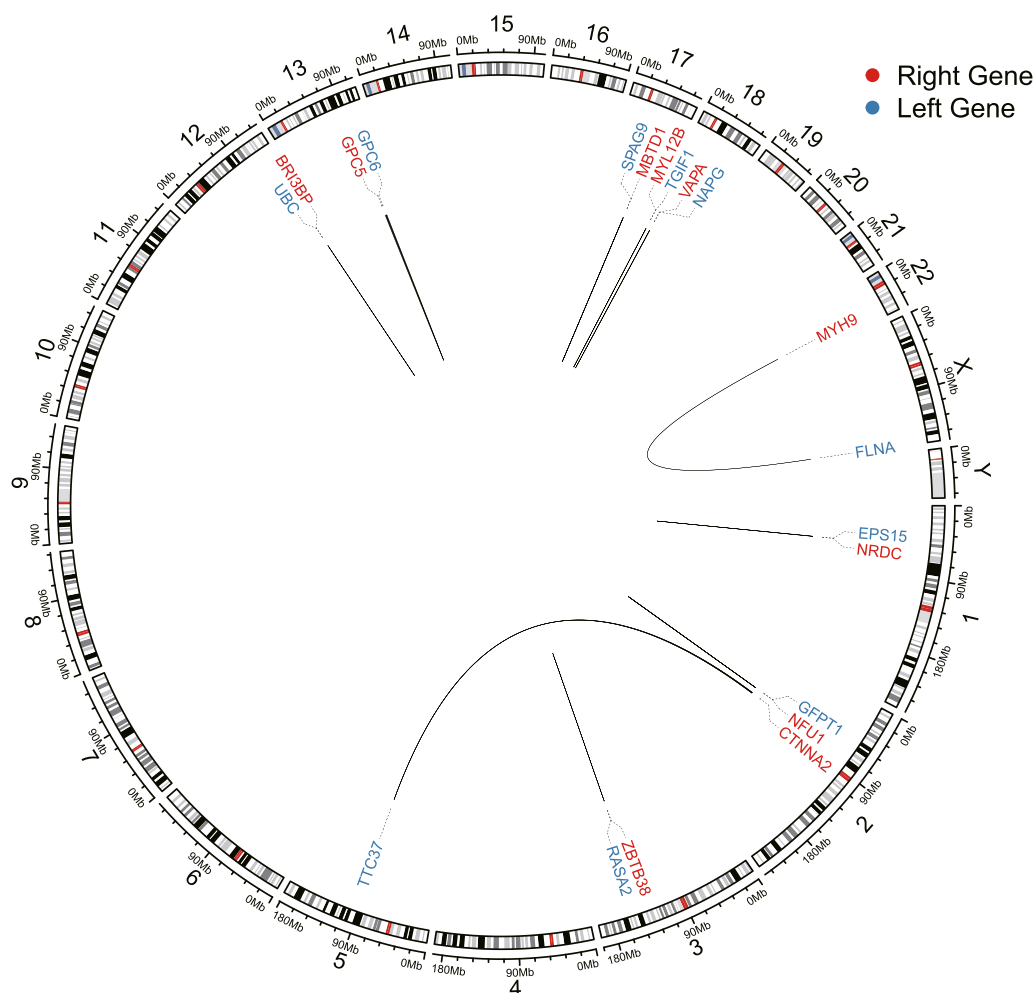


**Fig. 7. Examples of Omics-Omics (transregulation) correlation analysis in the “Omics-Omics (transregulation)” module.** By clicking the “START” button, this module will present a trans-regulatory correlation plot based on users’ input omics-gene lists, in which the samples are ordered ascendingly by the omics level of reference, white colour is used to present NA values, and the statistical results (for reference with continuous settings (A), black colour represents that the absolute correlation coefficients are more than user-defined cut-off; when the data type of reference is discrete (B), black colour represents that *P* or FDR values are less than 0.05) are shown on the right-hand side. Except for DNA methylation profiles, the omics levels of variables with continuous settings illustrated in the figure were normalized by deviation standardization.

genomes. In various instances, gene fusions play an important role in oncogenesis. In this module, a circos plot was used to visualize the comprehensive gene fusion landscape of the user-defined cell line in the “Fusions” module of IMOPAC. For example, the fusion landscape of HUH28, one of the bile duct cancer cell lines, was visualized based on the DepMap dataset in Fig. 8.

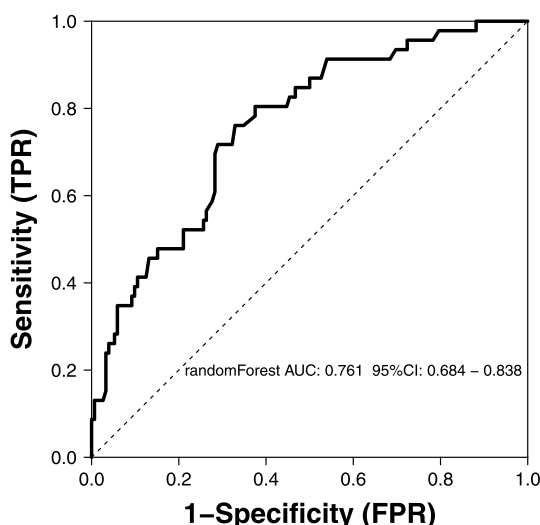
### 3.9. Prediction of drug response analysis

In the “Machine Learning in Predicting Drug Response” module of IMOPAC, users could investigate the performance of multiomics signature, including mRNA, DNA methylation, CNV to predict the response of user-defined drug using ten machine learning algorithms with 5-fold stratified cross validation, which was visualized by AUC curve in GDSC dataset. For example, the ability of hypoxia signature in



**Fig. 8.** The circos link plot is provided in the “Fusions” module based on users’ custom input cell line.





**Fig. 9. Examples of Prediction of Drug response analysis in the “Machine Learning in Predicting Drug Response” module.** By clicking the “START” button, this module will present an AUC curve based on the specific machine learning algorithm (Random forest), which will visualize the performance of predicting the Nilotinib response by hypoxia signature. (Hypoxia signature [14] is based on 15-gene expression signature (ACOT7, ADM, ALDOA, CDKN3, ENO1, LDHA, MIF, MRPS17, NDRG1, P4HA1, PGAM1, SLC2A1, TPI1, TUBB6 and VEGFA)).

prediction of Nilotinib is visualized in Fig. 9 using random forest algorithm.

### 3.10. Documentation

The tutorial and examples of IMOPAC are available and can be accessed by clicking the ‘HELP’ link in the top right navigation bar. The tutorial contains the description of each module function, the query procedure and the introduction of parameters in each module as well as the results of each analysis. Meanwhile, all data involved in IMOPAC can be downloaded at the ‘DOWNLOAD’ page.

## 4. Discussion

IMOPAC is an interactive web application for interactive multiomics and pharmacological analyses of patient-derived cancer cell lines originating from three publicly available databases, including GDSC, DepMap, and NCI60, using the output of a standard processing pipeline for pharmacogenomics.

It was primarily developed to comprehensively characterize the landscape of pharmacogenomic interactions and easily visualize the pharmacogenomic profiles in preclinical cancer models. Our web server provides biologists, without the need for complex programming technology or expertise, with the ability to perform pharmacogenomics mining in the discovery of the potential biological mechanism of drug resistance or repurpose new anticancer therapeutics and identify clinically actionable biomarkers by the correlation analysis between drug sensitivity and multiomics alteration or pathway activity and allows users to easily identify which cell line models most closely resemble specific cancer subtypes and provide a better understanding of biological mechanisms by directly linking in vitro models with patient tumour samples with curated clinical data. What’s more, by utilizing the overlapping drug response data of the same cell lines in both the DepMap and GDSC databases and conducting correlation analysis, we have found that the two databases maintain good consistency (Fig. S1). In addition, users can also perform pancancer expression analysis to assess a comprehensive understanding of a given gene’s molecular pattern across multiple different cancer types in our portal. Because genes with

similar expression features have a tendency to possess related functions, it is frequently desirable to detect genes with expression similarities to a known gene, acting as a known tumour chemotherapy target. In doing so, cis- and trans-regulation correlation analysis can be performed in IMOPAC to identify multiomics molecules with homologous expression patterns. Because gene fusions play an important role in oncogenesis, the comprehensive gene fusion landscape of the user-defined cell line can also be visualized in our tool. Finally, the drug response prediction module in IMOPAC on the basis of GDSC database, which provides 10 different machine learning algorithms for predicting response to various drugs based on user-defined omics data and gene sets, presents its potential value in guiding anti-cancer drug selection. However, it is crucial to note that the predictive model may be affected by other factors, and it is necessary to conduct further validation on an independent dataset in the future analysis.

At the same time, the custom drawing parameters of IMOPAC allow users great freedom to customize the visual output, for example, by selecting the type of omics, the method of correlation analysis, cluster algorithm, form of display for *P*-values, or ascertaining the cut-off of regression coefficient.

IMOPAC will be updated quarterly, with more analytical functions and interactive visualizations developed, as well as additional valuable cohort-based public datasets integrated in the future, which will collect multiomics and drug screens from patient-derived tumour xenografts and organoids. With these further developments, IMOPAC will complement other available tools to assist biological and clinical researchers in characterizing pharmacogenomics, capturing the genetic diversity of cancer and identifying drug response biomarkers across pancancer.

IMOPAC is a time-saving and intuitive web-based tool that provides users, especially biologists without background knowledge of computer programming, with a powerful tool for biomarker discovery, drug repurposing, and precision treatment. With continuous user feedback and further enhancement, IMOPAC has the potential to become an integral part of routine data analysis for experimental biologists.

### Ethics approval and consent to participate

Not applicable.

### Funding

Not applicable.

### CRediT authorship contribution statement

Ganxun Li, Dongyi Wan, and Junnan Liang: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing - Original Draft, Peng Zhu, Zeyang Ding, and Bixiang Zhang: Conceptualization, Funding Acquisition, Resources, Supervision, Writing - Review & Editing.

### Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

### Data availability

IMOPAC is free and open to all users and has no login requirement. All IMOPAC source code (back-end, front-end and data processing) can be downloaded from GitHub repository: <https://github.com/wolfgangk07/IMOPAC>.

### Acknowledgements

The authors would like to thank colleagues in the GDSC, DepMap and NCI60 projects. The authors also acknowledge the resources of GDSC,

DepMap, NCI60, HGNC, Cell Model Passports, PubChem and MSigDB.

#### Consent for publication

Not applicable.

#### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2023.07.023](https://doi.org/10.1016/j.csbj.2023.07.023).

#### References

- [1] Prasad V. Perspective: the precision-oncology illusion. *Nature* 2016;537(7619):S63.
- [2] Boehm JS, et al. Cancer research needs a better map. *Nature* 2021;589(7843):514–6.
- [3] Malone CF, et al. Selective modulation of a pan-essential protein as a therapeutic strategy in cancer. *Cancer Discov* 2021;11(9):2282–99.
- [4] Iorio F, et al. A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;166(3):740–54.
- [5] Yang W, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;41:D955–61 (Database issue).
- [6] Ghandi M, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019;569(7757):503–8.
- [7] Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483(7391):603–7.
- [8] Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;6(10):813–23.
- [9] Liu M, et al. Integrative epigenetic analysis reveals therapeutic targets to the DNA methyltransferase inhibitor guadecitabine (SGI-110) in hepatocellular carcinoma. *Hepatology* 2018;68(4):1412–28.
- [10] Chai P, et al. Genetic alteration, RNA expression, and DNA methylation profiling of coronavirus disease 2019 (COVID-19) receptor ACE2 in malignancies: a pan-cancer analysis. *J Hematol Oncol* 2020;13(1):43.
- [11] Hu X, Zhang Z. Understanding the genetic mechanisms of cancer drug resistance using genomic approaches. *Trends Genet* 2016;32(2):127–37.
- [12] Zhou J, et al. Pan-ERBB kinase inhibition augments CDK4/6 inhibitor efficacy in oesophageal squamous cell carcinoma. *Gut* 2022;71(4):665–75.
- [13] Li G, et al. Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma. *EBioMedicine* 2019;47:128–41.
- [14] Ye Y, et al. Characterization of hypoxia-associated molecular features to aid hypoxia-targeted therapy. *Nat Metab* 2019;1(4):431–44.
- [15] Gillani R, et al. Gene fusions create partner and collateral dependencies essential to cancer cell survival. *Cancer Res* 2021;81(15):3971–84.
- [16] Roosen M, et al. The oncogenic fusion landscape in pediatric CNS neoplasms. *Acta Neuropathol* 2022;143(4):427–51.
- [17] Jia P, et al. Deep generative neural network for accurate drug response imputation. *Nat Commun* 2021;12(1):1740.
- [18] Luna A, et al. CellMiner Cross-Database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res* 2021;49(D1):D1083–93.
- [19] Feizi N, et al. PharmacoDB 2.0: improving scalability and transparency of in vitro pharmacogenomics analysis. *Nucleic Acids Res* 2022;50(D1):D1348–57.
- [20] Dong Q, et al. RNAactDrug: a comprehensive database of RNAs associated with drug sensitivity from multi-omics data. *Brief Bioinform* 2020;21(6):2167–74.
- [21] Liberzon A, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1(6):417–25.
- [22] Yang X, et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 2014;26(4):577–90.
- [23] van der Meer D, et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res* 2019;47(D1):D923–9.
- [24] Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res* 2021;49(D1):D939–46.
- [25] Kim S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res* 2021;49(D1):D1388–95.
- [26] Leda AR, et al. Insights about minority HIV-1 strains in transmitted drug resistance mutation dynamics and disease progression. *J Antimicrob Chemother* 2018;73(7):1930–4.
- [27] Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma* 2013;14:7.
- [28] Gu Z, et al. circlize Implements and enhances circular visualization in R. *Bioinformatics* 2014;30(19):2811–2.