BMC Bioinformatics

SOFTWARE

Open Access

MultiOmicsXplorer, a tool to browse, access and analyse multi-omics data

Eleonora Meo[1†], Veronica Lombardi[2†], Veronica Venafra[2], Valerio Licursi[3], Francesca Sacco[1*] and Livia Perfetto[2*]

†Eleonora Meo and Veronica Lombardi have contributed equally to this work.

*Correspondence:
Francesca Sacco
francesca.sacco@uniroma2.it
Livia Perfetto
livia.perfetto@uniroma1.it
[1]Department of Biology, University of Rome Tor Vergata, Rome, Italy
[2]Department of Biology and Biotechnologies 'Charles Darwin', Laboratory affiliated to Istituto Pasteur Italia-Fondazione Cenci Bolognetti, Sapienza University of Rome, Rome, Italy
[3]Institute of Molecular Biology and Pathology (IBPM), National Research Council (CNR), Rome, Italy

Abstract

Background Personalized and precision medicine aim to identify predictive biomarkers from patient-specific proteogenomic profiles and to uncover tailored therapeutic strategies by targeting deregulated proteins driving disease phenotypes.

Methods To address these challenges, we developed *MultiOmicsXplorer*, a freely available and interactive R-based Shiny application designed to facilitate the exploration and analysis of multi-omics cancer datasets with a user-friendly interface. At its core, *MultiOmicsXplorer* builds on our recently developed SignalingProfiler pipeline to extract protein activities from proteogenomic data, thereby reducing data complexity and dimensionality while enabling mechanistic hypothesis generation.

Results The application integrates two core functionalities. The first, *OncoXplorer*, enables the systematic inference and comparison of protein activities from 1492 samples corresponding to approximately 1000 patients across ten cancer types, leveraging harmonized datasets from the CPTAC portal to provide a functional and mechanistic interpretation of multi-layered data. Importantly, beyond activity estimation, it allows for comparative analysis of transcriptomic, proteomic, and phosphoproteomic data. The second module, *Extract Protein Activity from Your Data*, enables users to infer the activity of kinases, phosphatases, and transcription factors from custom multi-modal datasets.

Conclusions Overall, *MultiOmicsXplorer* facilitates the exploration and interpretation of CPTAC multi-omics cancer datasets, supporting comparative analyses and protein activity inference within a unified and user-friendly environment.

Keywords Proteogenomic data, Protein activity estimation, Personalized medicine, Biomarkers, Pan cancer

Background

Recent technological advancements have made it possible to measure the concentrations of thousands of transcripts and proteins—including their post-translational modifications—in patient-derived cancer specimens [1–4]. In principle, these large-scale omics datasets offer an unprecedented opportunity to discover prognostic biomarkers and design novel personalized therapeutic strategies [5]. Public repositories such as the Clinical Proteomic Tumor Analysis Consortium (CPTAC) data portal, which hosts omics

data from hundreds of cancer patients, make these extensive datasets freely accessible to the scientific community [6]. However, leveraging these datasets to identify prognostic biomarkers and potential therapies remains challenging: while analyte-level changes provide quantitative information, they do not directly reveal the functional state of signaling regulators that ultimately drive cellular phenotypes. To overcome this limitation, several computational tools have been developed to generate mechanistic hypotheses from large datasets [7–10]. Among these tools is *SignalingProfiler* 2.0 [11] which aims to address this challenge by identifying deregulated signaling molecules from multi-omics datasets. Specifically, it infers the functional state of kinases, phosphatases, and transcription factors from their downstream targets. However, the application of SignalingProfiler demands advanced computational skills, thereby limiting its accessibility to researchers without a strong computational background.

Here, we introduce *MultiOmicsXplorer*, a free and user-friendly computational framework designed to facilitate the extraction of protein activity from multi-omics datasets using SignalingProfiler, enabling researchers to move from high-dimensional analyte abundance profiles to functional regulatory insights. *MultiOmicsXplorer* enables the analysis of transcriptomic, proteomic, and phosphoproteomic data, inferring the activity of regulatory proteins in publicly-available CPTAC datasets as well as in user-defined experimental conditions. We believe *MultiOmicsXplorer* will substantially enhance the systematic analysis of large-scale omics datasets and facilitate hypothesis-driven discovery.

Implementation

CPTAC data analysis

Step 1—Access to CPTAC data (Supplementary Fig. 1A). As a first step, we used the Linkedomics web portal [12] (October 2025) to retrieve proteomic, phosphoproteomic, and transcriptomic datasets containing the abundance of transcripts, proteins, and phosphorylation sites [13], extracted from 1492 tumor and non-tumor tissues and covering ten cancer types registered in the CPTAC portal [6]: *Breast cancer (Brca)* [1], *Clear cell renal cell carcinoma (Ccrcc)* [14], *Colon adenocarcinoma (Coad)* [15], *Glioblastoma (Gbm)* [16], *Head and neck squamous cell carcinoma (Hnscc)* [17], *Lung squamous cell carcinoma (Lscc)* [18], *Lung adenocarcinoma (Luad)* [19], *Ovarian cancer (Ov)* [20], *Pancreatic ductal adenocarcinoma (Pdac)* [21], *Uterine corpus endometrial carcinoma (Ucec)* [22] (Supplementary Fig. 2).

Importantly, data retrieved from Linkedomics are already pre-harmonized as follows: (i) Transcriptomic data were generated from RNA-seq experiments processed through gene-level quantification using RSEM and upper-quartile normalization, with expression values reported as log2-transformed normalized abundances. (ii) (Phospho)proteomic data were generated using TMT-based mass spectrometry, log2-transformation, and normalization to a common reference channel within each TMT plex. Next, intensities were summarised at the gene-, protein- and site-level respectively, followed by outlier removal, batch effect correction and cross-cancer harmonization [13].

Step 2—Data manipulation (Supplementary Fig. 1B). We used the data manipulation strategy implemented in our recently developed PatientProfiler tool [23] to ensure a coherent distribution of the data, achieve standardization across all tumor types, and maintain comparability of the analyzed data. Key steps in this data manipulation include

"analyte" updating, imputation, filtering, and sample-wise Z-score computation. The term "analyte" here represents gene- and protein-level measurements in transcriptomics and proteomics, and phosphorylation site measurements in phosphoproteomics. The result of the manipulation is a collection of datasets that are formatted for downstream analyses.

Update. For the transcriptomic dataset, the updating process included a check on the existence of the provided primary gene name.

For the phosphoproteomic datasets, duplicate entries sharing the same gene name and/or UniProtKB identifier, and phosphorylation site were aggregated at the sample level by averaging their expression values. Furthermore, if a modification site was associated with more than one UniProtKB ID or gene name, these identifiers were merged into a single entry and separated by semicolons.

Missing values management. We used the multiple imputation by chained equations (MICE) method [24, 25] to impute missing values in the proteomic and phosphoproteomic datasets. Before imputation, features (rows) with more than 80% missing values were removed. MICE was performed considering the Tumor–non-Tumor status as a covariate, generating five imputations for each missing value; the final value was obtained by computing the median across the five estimates.

For the transcriptomic data, NAs were identified only in three datasets: *breast cancer*, *lung squamous cell carcinoma* and *lung adenocarcinoma*. In these datasets, NAs were replaced with 0 values. Subsequently, for every transcriptomic dataset, we excluded rows (analytes) containing more than 80% of 0 values.

Filter. We applied a filter to retain only analytes common to at least two tumor datasets. This approach allowed us to reduce potential noise from analytes that may be tumor-specific or unrelated, focusing on those that have a higher probability of reflecting shared mechanisms or biomarkers across tumors.

Z-score calculation. To make the analysis independent of the number of samples, we calculated the sample-wise Z-score (within each patient). This approach allowed us to standardize the data for each individual patient. In addition, this normalization method ensures that comparisons are based on the relative deviation of each analyte from the patient's median value, neutralizing variations in data due to differences between patients. For each analyte in each sample, Z-score values falling outside the 95% confidence interval (Z-score < -1.96 or $> +1.96$) were considered significantly deregulated (p-value < 0.05), representing deviations from the sample median, rather than direct evidence of functional dysregulation. Supplementary Fig. 3 summarizes the distribution of significant analytes for each patient and each omic layer.

Step 3—Protein activity inference (Supplementary Fig. 1C). We used the SignalingProfiler 2.0 tool [11] to compute protein activities. Briefly, SignalingProfiler allows the prediction of kinase and phosphatase activity from the abundance of their substrates in the phosphoproteomics and the prediction of transcription factor activity from the abundance of the target genes in the transcriptomics.

Specifically, we used transcription factor enrichment analysis (TFEA), kinase substrate enrichment analysis (KSEA) and footprint-based analysis to infer the activity of key signal transduction proteins.

The analyses involving TFEA, KSEA and PhosphoScore were conducted using default parameters.

Supplementary Fig. 4 shows a summary of the count of total proteins whose activity was estimated for each tumor, organized by molecular function: kinases, transcription factors, phosphatases and other phosphorylated proteins.

MultiOmicsXplorer development

For the development of the *MultiOmicsXplorer* we used the R "shiny" package (version 1.10.0, https://shiny.posit.co/). The source code of the application is available at https://github.com/SaccoPerfettoLab/MultiOmicsXplorer.

The application is divided into two main parts:

- *"OncoXplorer—CPTAC data analysis".* This section allows users to explore and compare pre-estimated protein activities from the multi-omics data from the CPTAC portal using the SignalingProfiler 2.0 pipeline; in addition, this section offers the possibility to directly compare analyte abundances from the same datasets.
- *"Extract Protein Activity from your data".* This section allows users to infer protein activities from their phosphoproteomics, proteomics, and transcriptomics files using the SignalingProfiler 2.0 pipeline.

The UI (user interface) was built using the "shinydashboard" package to organize the different tab panels and boxes for input selection and for the generation of plots and tables.

OncoXplorer—CPTAC data analysis. To generate comparisons, we formatted and adapted previously harmonized datasets for each tumor type. Phosphoproteomic, proteomic and transcriptomic data were merged with corresponding clinical information (Stage and Patient_ID). In the same way, we formatted protein activity data obtained in the last step of processing and merged them with clinical data.

To visualize the levels of the analyte abundance and protein activity, we implemented two core functions that process the user-provided inputs in their respective analysis panels (i.e. *generate_plot_abundance* for *"Explore Analyte Abundance"* and *generate_plot_protein_activity* for *"Explore Protein Activity"*).

The arguments for the generation of the comparison plots are:

- The "omic" layer—phosphoproteomics, proteomics or transcriptomics—for the analyte abundance comparison;
- The "analyte" parameter, which depends on the omic layer selected for the analyte abundance comparison and the protein for the "*Explore Protein Activity*" panel;
- The analyte filter type, which allows users to choose to include all the samples or to restrict the analysis to those samples in which the analyte of interest exhibits a statistically significant Z-score ($|Z| > 1.96$, corresponding to $p < 0.05$);
- The "comparison type" parameter, followed by the "tumors" parameter, that determines the grouping of samples for differential analysis, to have a comparison between up to 10 tumors ("*Tumor vs. Tumor*") or between tumor samples and the non-tumoral tissues ("*Tumor vs. Non-tumor*"), depending on the Patient_ID;
- The "stages" parameter allows users to filter samples based on tumor stage (from 1 to 4). Alternatively, the "all" option aggregates all available stages by computing an average per tumor type.
- The "show distribution" and "show significance" parameters, when selected, show, respectively, the distribution of each sample in the boxplot and the Wilcoxon test (calculated using the *stat_compare_means* function using the "*wilcox.test*" method

from *ggpubr* package version 0.6.0) for each combination of tumors and stages. If one tumor and multiple stages are selected, the Wilcoxon test will be realized among the stages for the same tumor. However, if multiple tumors are selected, the test will be performed among tumors for each selected stage. The Wilcoxon test is also calculated in "*Tumor vs. Non-tumor*" comparisons.

The query returns a box plot generated with the *ggplot* function (ggplot2 package, version 3.5.2) and a table containing all formatted data is also automatically generated. Both the plot and the table can be downloaded using the *downloadButton* function of the *shiny* package.

As an additional feature, with the selection of the "Phosphoproteomics" parameter to generate a comparison box plot for a specific phosphorylation site, we implemented the "*query_ptm_residue*" function that performs an automated query through the RESTful API made available by the curators of the SIGNOR database [26] to verify the presence of the selected site. Specifically, the function uses the *get* method from the *httr* package (version 1.4.7) to send an HTTP request to SIGNOR's residueSearch.php endpoint, using the UniProt ID of the protein of interest and the phosphorylated residue as parameters.

If the site is present in the database, the system automatically generates a structured URL linking to the dedicated page of the PhosphoSIGNOR platform (*entity.php?role=residue&ID=UniProtID_Protein_Residue*).

Extract protein activity from your data. This functionality was implemented by generating a wrapper function (*perform_analysis*) on the SignalingProfiler package. We parsed the inputs given in the user interface to obtain the arguments of SignalingProfiler functions.

To note, here users can upload multiple files: data can derive from multiple samples, each identified by a user-defined identifier (e.g., the patient's ID), or from multiple omic layers, where transcriptomic, proteomic, or phosphoproteomic datasets are indicated by the 'T', 'P' or 'Ph' prefix, respectively.

The first step, to run the SignalingProfiler analysis, is the upload of the file(s), making sure they are named according to the instructions provided above the upload box. To ensure flexibility in data input, we created a function (*read_data*) that handles the uploaded files by recognizing and reading the file format, supporting '.csv', '.tsv' and '.xlsx' extensions.

Another function, *check_columns*, verifies that each omic dataframe contains the required columns to perform the analysis. If any of the required columns are missing, the function returns a descriptive error message, which is displayed in the user interface.

Subsequently, users are required to specify the organism of interest (*human* or *mouse*) and may optionally customize a set of parameters that correspond to the arguments of the *perform_analysis* function. During the analysis, the results are temporarily stored in a unique RDS file created for each individual analysis. The file is used exclusively during the analysis process and is automatically deleted once the analysis is completed, ensuring that no data overlap occurs between different analyses.

The analyses carried out by the SignalingProfiler pipeline include transcription factor enrichment (TFEA), kinase-substrate enrichment (KSEA) and phosphoscore computation. The organism choice determines the phosphoscore calculation, involving BLASTP-based homology mapping to human data, when the organism selected is "mouse".

All the results from TFEA/KSEA are combined with the phosphoscore and the resulting output tables appear:

- "*Protein Activity Inference Results*", containing the inferred activity scores for kinases, phosphatases, transcription factors and other proteins for each sample;
- "*Summary Table*", which summarizes the count of the proteins categorized according to the molecular function, for each sample.
- "*Regulons Evidence Table*" which returns, for each inferred protein, downstream analytes driving the inference (proteins, phosphosites, and/or genes); their measured abundance values in the uploaded dataset; the regulatory sign of each interaction (activatory vs. inhibitory); the analysis method contributing to the score (KSEA, TFEA, or PhosphoScore).

The resulting tables can also be downloaded, and the "*Protein Activity Inference Results*" table can be used in the "*Plot Results*" section. For this section of *MultiOmicsXplorer*, the function *create_top_proteins_plot* was developed to generate box plots representing, by default, the top 15 up-regulated and 15 down-regulated proteins, divided by molecular function, generated with the *ggplot* function (*ggplot2* package, version 3.5.2). The plotting function is interactive, users can choose to modify the number of proteins to display, to change the plot size (height and width) and select, via a *radioButton*, to compute the average across all samples or select a specific sample from a dropdown menu.

Exploring CDK1 regulation with OncoXplorer

To display the protein activity of CDK1, and WEE1 in different tumor types, we followed the procedure below:

(1) We selected within the *Explore Protein activity* tab the specific gene name;
(2) We selected the "Tumor vs. Tumor" comparison. We then selected the tumors of interest from the drop-down menu: Brca, Ccrcc, Gbm, Hnscc, Ov and Pdac;
(3) We choose the "all" option to visualize the general trend of regulation, without specifying the stage.
(4) We selected the 'Show significance (Wilcoxon Test)'.

To display the protein activity of CDK1 and WEE1 in different tumor stages of Pdac, we followed the following procedure:

(1) We selected within the *Explore Protein activity* tab the specific gene name;
(2) We selected the "Tumor vs. Tumor" comparison. We then selected Pdac;
(3) We choose the 1, 2, 3, 4 option to visualize the general trend of regulation, across the stages.

User testing

In the final stages of development, we conducted a user testing phase for *MultiOmicsXplorer*, by recruiting ten users who had never used the application.

Results

MultiOmicsXplorer is accessible at https://perfettolab.bio.uniroma1.it/MultiOmicsXplorer/.

At the very core of *MultiOmicsXplorer* is the notion that we can extract protein activities from multi-omics dataset in order to reduce data complexity and dimensionality and extract mechanistic information [27]. As shown in Fig. 1, the application is composed of two main parts, both employing of our recently developed SignalingProfiler tool to infer protein activities: the "*OncoXplorer—CPTAC data analysis*" section, which enables users to compare protein activities, as well as gene-, protein- and phosphosite-level abundances, across tumor types and tumoral stages leveraging pre-harmonized data from the CPTAC portal; and the "*Extract protein activity from your data*" section, which provides a user-friendly graphical interface allowing users to infer protein activities from custom datasets (Fig. 2). Importantly, this capability extends the application of *MultiOmicsXplorer* to a broader set of samples, not limited to the cancer context.

A detailed tutorial and toy datasets are available on the website to facilitate understanding of the required data format and the functionality of the tool. Also, sample datasets are provided to assist in the preparation of files.

OncoXplorer—CPTAC data analysis

The *OncoXplorer* section of the application organizes and provides access to patient-specific information across ten different tissue types, derived from the CPTAC portal (Fig. 1B, Supplementary Fig. 1). It presents two subsections: (1) the *"Explore Analyte Abundance"* which allows users to display and compare the abundance of single analytes, chosen from the three "omic" levels—phosphoproteomics, proteomics, and transcriptomics—among tumor types and/or tumoral stages; and (2) the "*Explore Protein Activity*" subsection, which shows differences in SignalingProfiler-estimated protein activities within the same comparison groups. Note that the term "analyte" is hereby used to refer to the quantified molecules within each omic level: it represents gene- and protein-level measurements in transcriptomics and proteomics, and phosphorylation site measurements in phosphoproteomics.

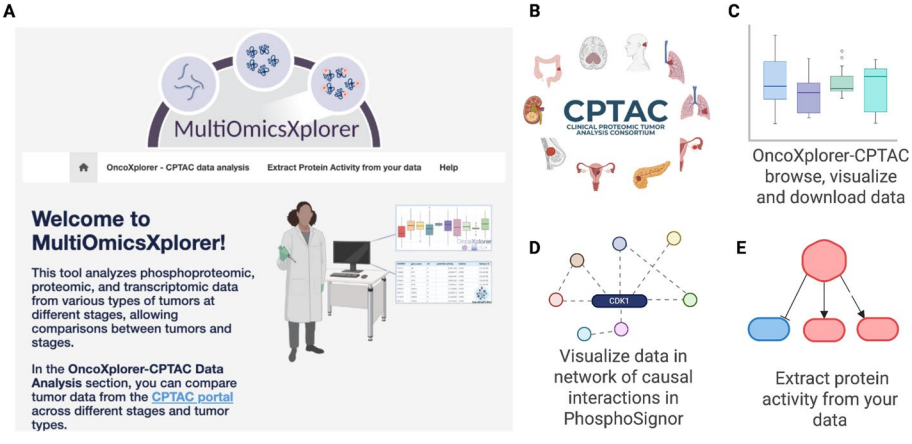

Fig. 1 Summary of the features of *MultiOmicsXplorer*. **A** *MultiOmicsXplorer*, the website homepage. **B** A scheme of the ten tumor types selected from the CPTAC portal integrated in the application. **C** In the *OncoXplorer* section it is possible to compare analyte abundance and protein activity across tumor types and tumoral stages, leveraging pre-harmonized data. **D** *OncoXplorer* gives users direct access to the network of causal interactions (in the SIGNOR resource) surrounding a specific phosphosite analyte. **E** In the *Extract protein activity from your data* section, users can estimate protein activities from private uploaded data. Figure created with BioRender.com

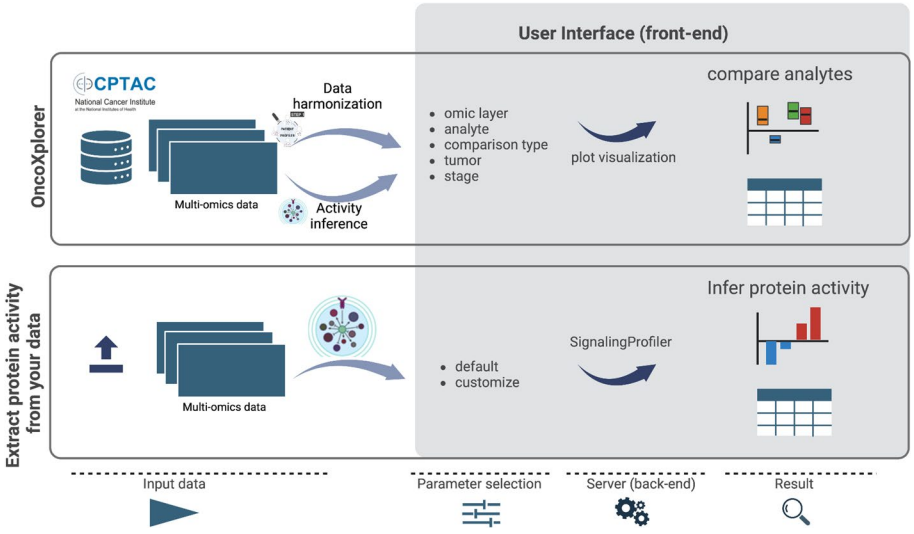

Fig. 2 Structure of the *MultiOmicsXplorer* application. The structure of the application is made of two main sections: The first takes as input CPTAC data [13] (*OncoXplorer—CPTAC data analysis* section) to obtain box plots with comparisons of analyte abundances and SignalingProfiler-derived protein activities. The second accepts as input user-defined datasets (*Extract protein activity from your data*) to obtain tables with activities of kinases/transcription factors/phosphatases/other proteins inferred by the SignalingProfiler tool [11]. Figure created with BioRender.com

Very briefly, the CPTAC consortium recently launched an initiative to collect proteogenomic information from individual patient samples; to standardize data collection, tumor purity selection and analysis procedures; and to release this body of data in a centralized resource [13]. This effort enabled us to access a harmonized proteogenomic dataset encompassing ten distinct cancer types from 1492 samples, including both tumoral and non-tumoral tissues—except for breast and glioblastoma cancer datasets, which contain only tumoral samples (Supplementary Fig. 2). We applied data-processing procedures which included meticulous data cleaning and parsing of more than 200,000 analytes (Supplementary Fig. 2). The cleaning process was followed by missing-data filtering and imputation strategies and standardization by Z-score calculation. Importantly, we next employed the SignalingProfiler pipeline to infer protein activity from the multi-omics layers (see Implementation section for additional details) (Supplementary Fig. 1).

The *"Explore Analyte Abundance"* panel enables users to select an omic layer from the following options: transcriptomics, proteomics, or phosphoproteomics. Based on this selection, a drop-down menu containing the list of available analytes is loaded. After selecting the analyte of interest, users can set alternative parameters and visualize/download the result as a box plot (Fig. 1C) and as a table. Here we summarize a list of functionalities offered.

(1) *Sample filtering*: selecting "Significant analytes" restricts the comparison to analytes with significant Z-scores ($Z < -1.96$ or $Z > +1.96$). Alternatively, selecting "All analytes" includes all analytes in the comparison (Supplementary Fig. 3).

(2) *Comparison type*: users can choose to compare analyte abundance across up to ten different cancer types ("Tumor vs. Tumor") or between malignant and non-tumoral samples in the same tissue ("Tumor vs. Non-tumor").

(3) *Tumor stage selection*: users can specify the tumor stage to be considered (Stage 1–4). Alternatively, the "all" option is available.

(4) *Plot customization*: users can opt to visualize individual data points, include statistical significance in the resulting plot (Wilcoxon test) and interactively modify the plot size (width and height).

(5) *Phosphosite functional context*: the selection of the "Phosphoproteomics" parameter runs an automated query to the SIGNOR database [26] to check for the presence of the selected phosphorylation site in the resource. This link gives users direct access to the network of causal interactions surrounding the phosphorylated site (e.g. upstream kinases, and the regulatory effect on the target protein), providing a functional context of the post-translational modification within cellular signaling pathways (Fig. 1D).

To date, the "*Explore Analyte Abundance*" allows users to visualize 202,809 analytes (60,669 transcripts, 15,100 proteins, and 127,040 phosphosites) across ten cancer types and four stages. Importantly, these abundance measurements provide a comprehensive quantitative overview of molecular alterations in cancer.

The "*Explore Protein Activity*" complements this analyte-centered perspective and shifts the focus from molecular abundance to functional regulation, within the same comparison groups described. In brief, it allows users to display differences in estimated protein activities derived from multi-omics data processed with the SignalingProfiler software (see Implementation section). As such, it is similar in structure to the *"Explore Analyte Abundance"* section: it includes a drop-down menu containing all the proteins for which activity has been calculated using SignalingProfiler in our data analysis pipeline (Supplementary Fig. 4). Users can select a protein and choose between two comparison options: "Tumor vs. Tumor" or "Tumor vs. Non-tumor". Depending on the selection, the interface adjusts to enable the specific comparison of tumors or clinical stages. Additionally, users can opt to display individual data points and results from the Wilcoxon significance test.

All the queries in the *OncoXplorer* section will return a customizable box plot (Supplementary Fig. 6) and a table, available for download.

Extract protein activity from your data

The second main section of *MultiOmicsXplorer* allows users to apply SignalingProfiler to their own multi-modal datasets and to derive the activity of transcription factors from transcriptomics and kinases, phosphatases, and other proteins from (phospho)proteomics (Fig. 1E). The "*About SignalingProfiler*" panel provides a brief overview of the method and its applications, with references to the original publication.

The "*Run Analysis*" panel is the operational interface that allows users to upload and analyze data. The system supports multiple tabular file formats and accepts transcriptomic, proteomic, and phosphoproteomic datasets alone or in combination. The process is highly customizable thanks to the advanced parameter panel, which allows users to tune key parameters of the SignalingProfiler pipeline (see Venafra et al. 2024 for details [11]). The analysis returns two main tables: the first reports the proteins whose inferred activity is significantly altered in the analyzed dataset; the second provides, for each estimated protein, a detailed breakdown of the evidence supporting the inference, such

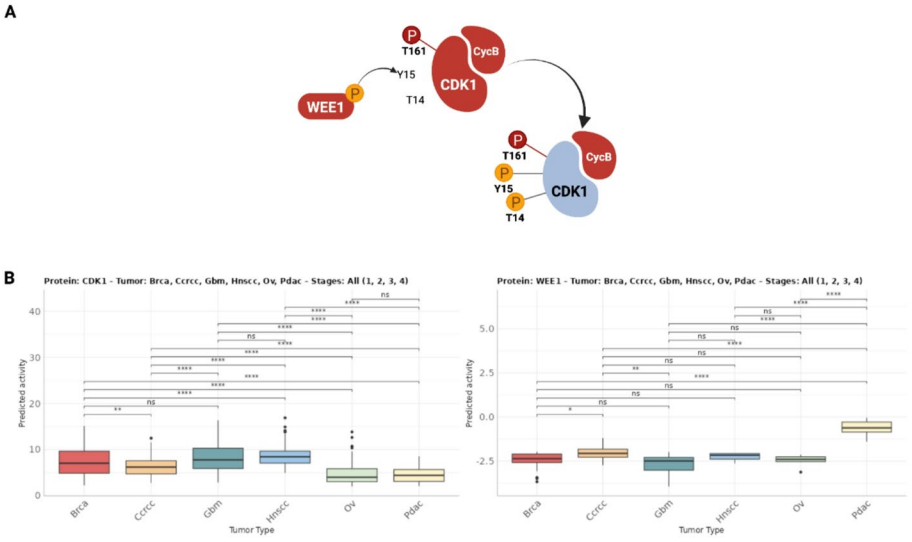

Fig. 3 Regulation of CDK1 by phosphorylation. **A** CDK1's activity is mediated at multiple levels: first, by its expression; second, by its interaction with the CyclinB1 (CCNB1); third, at the phosphorylation level. In particular, it can be phosphorylated both at the activatory Thr161 and/or at two inhibitory residues (Thr14 and Tyr15 by WEE1) [29] **B** Comparison box plot generated by *OncoXplorer—CPTAC data analysis*, displaying the activity level of CDK1 (left) and WEE1 (right) kinases in samples of patients affected by Brca (red), Ccrcc (orange), Gbm (blue), Hnscc (light blue), Ov (green), Pdac (yellow). Figure created with BioRender.com

as (i) the 'sources of information' (e.g., target phosphosites and genes in the regulons); (ii) their measured abundance values in the uploaded dataset; and (iii) their regulatory role (activatory vs. inhibitory). Users can download the tables and summary results for further analysis and to visualize the plot in the "*Plot Results*" subsection. The plot displays, by default, the top 15 proteins that have been found to be up- and down-regulated according to their predicted activity. The plotting function is interactive: users can choose to modify the number of proteins to display or change the plot size (height and width). Finally, displayed proteins are grouped by molecular function: transcription factors, kinases, phosphatases, and other proteins (Supplementary Fig. 6).

Use case: Exploring CDK1 regulation with OncoXplorer

As a proof of principle, we set out to use the *OncoXplorer* section of *MultiOmicsXplorer* to compare the activity status of the cyclin-dependent kinase 1 (CDK1) and of its inhibitory kinase WEE1 in different tumor types. Briefly, CDK1 is a cyclin-dependent kinase, and its activity is essential for cells to enter the M phase of the cell cycle [28], whereas WEE1 negatively regulates CDK1 activity by phosphorylating its inhibitory residue Tyr15 [29] (Fig. 3A).

Using the *"OncoXplorer—CPTAC data analysis"* tool, we compared the predicted activity of CDK1 and WEE1 across six tumor types (Brca, Ccrcc, Gbm, Hnscc, Ov, Pdac) (Fig. 3B). The analysis revealed that CDK1 activity is significantly higher in breast cancer (Brca), glioblastoma (Gbm), and head and neck squamous cell carcinoma (Hnscc), while it is markedly lower in ovarian (Ov) and pancreatic ductal adenocarcinoma (Pdac) samples.

Conversely, WEE1 displays an opposite trend, showing lower activity in Brca, Gbm, and Ov, but elevated activity in Pdac. This anticorrelation between CDK1's and WEE1's

activity (Pearson's $r = -0.65$, $p < 2.4e{-}19$) reflects the expected functional antagonism between these kinases.

To further dissect this relationship, we explored the activity patterns of both kinases across the four stages of Pdac (Supplementary Fig. 5).

The stage-specific analysis revealed a gradual increase in CDK1 activity from stage 1 to 4, with WEE1 displaying the opposite behaviour (Supplementary Fig. 5). Although not statistically significant, these results indicate that CDK1 may progressively escape WEE1-driven inhibition during tumor progression.

Overall, this use case demonstrates how *MultiOmicsXplorer* enables mechanistic exploration of kinase activity across tumors and disease stages, with crucial implications for biomarker discovery and therapy assignment.

Conclusions

MultiOmicsXplorer is an interactive R-based Shiny application, freely accessible at https://perfettolab.bio.uniroma1.it/MultiOmicsXplorer/. *MultiOmicsXplorer* was developed to facilitate access to cancer datasets and enhance the analysis of multi-modal data for scientists with limited programming experience.

Firstly, in the "*OncoXplorer—CPTAC data analysis*" section, *MultiOmicsXplorer* enables users to compare the abundance of more than 202,000 analytes (transcripts, proteins, and phosphosites) derived from more than 1492 samples corresponding to approximately 1000 patients, associated with ten distinct cancer types. Analytes and protein abundances can be compared across tumor types and tumoral stages, leveraging data from the CPTAC portal. As an example, we used the tool to derive the activation state of CDK1 and its WEE1 antagonist in six different tumor types, demonstrating that *MultiOmicsXplorer* allows the formulation of mechanistic hypotheses. Secondly, the "*Extract protein activity from your data*" section allows users to infer the activity of signaling proteins such as kinases, phosphatases, and transcription factors from private datasets, complementing the analysis with a user-friendly graphical interface.

In summary, *MultiOmicsXplorer* is an easy-to-use web application designed to facilitate exploration of multi-omics datasets from the CPTAC consortium as well as to support comparative analyses and protein activity inference. Despite this, *MultiOmicsXplorer* presents limitations.

A first limitation of our study concerns the handling of potential batch effects inherent to the integration of heterogeneous datasets within *OncoXplorer—CPTAC data analysis*. To alleviate this issue and to allow for pan-cancer analysis, we exploited a post-harmonized dataset made available by the CPTAC consortium [6]. In addition, we applied a series of preprocessing steps, including the exclusion of cohort-specific analytes, removal of features with an excess of missing data, and within-sample z-score transformation. However, it is important to stress that, despite these efforts, residual technical variability across cohorts cannot be fully ruled out and should be considered when comparing analytes in different cancer types. A second limitation, in integrating information from the CPTAC portal, concerns the reduced use of clinical data: currently, comparisons across tumors are limited to Stage information only, and a comprehensive clinical–molecular integrative analysis is not yet supported. Another important limitation to take into consideration is that *MultiOmicsXplorer* adopts sample-wise z-scoring to identify statistically deregulated analytes. This choice was determined by the absence of matched

control samples and of multiple replicates. Although this approach has been employed in similar contexts [1] it is crucial to point out that this provides no direct evidence of functional dysregulation of the analytes.

Regarding the "*Extract protein activity from your data*", a limitation to consider is represented by the incomplete coverage of the prior knowledge space in public resources. As a matter of fact, presently *MultiOmicsXplorer* has the potential to estimate activity for approx. 7500 proteins (< 40% of the human proteome) [11]. To fill this gap, we have already implemented regular updates of the prior-knowledge information within SignalingProfiler [11].

Availability and requirements

Project name: MultiOmicsXplorer.

Project home page: https://perfettolab.bio.uniroma1.it/MultiOmicsXplorer/.

Project repository: https://github.com/SaccoPerfettoLab/MultiOmicsXplorer.

Operating system(s): Platform independent (web-based application).

Programming language: R/Shiny.

Other requirements: For local use: R 4.2.0 or higher, Shiny package, SignalingProfiler package (https://github.com/SaccoPerfettoLab/SignalingProfiler).

License: MIT License.

Any restrictions to use by non-academics: none.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-026-06460-w.

Supplementary Material 1

Acknowledgements

We would like to thank Dr. Andrea Petretto, Flavio Di Florio, Miriana Santacroce, and Valentina Marchionni for suggestions and user testing. Dr. Riccardo Bertini for the technical support.

Author contributions

LP and FS Conceived and supervised the project; EM and VLO Developed and implemented the web application; EM and VLO Contributed to the bioinformatics requirements analysis and workflow design; EM, VLO, VV and VLI Performed data analysis and testing; EM, VLO, LP and FS Contributed to the validation of the application and interpretation of the results; EM, LP and FS Wrote the first draft of the manuscript; VV and VLO Participated in the critical revision for scientific and technical content; VLI Managed system integration and deployment; EM Designed the user interface and user experience; LP and FS Coordinated team activities and the manuscript submission process; LP and FS funding acquisition. All authors reviewed the manuscript.

Funding

This research was funded by the Italian Association for Cancer Research (AIRC) with a grant to L.P. (My First AIRC Grant "Giancarlo Delli Colli" n. 28858) and a grant to F.S. (Startup Grant n. 23099). Also, L.P. and F.S. are supported by a joint PRIN 2022 PNRR grant (n. P2022JRETW), funded by the European Union—NextGenerationEU. L.P. is also supported by an Istituto Pasteur—Fondazione Cenci-Bolognetti under 40 grant and by a Sapienza Grant. V.V. is supported by PON-MUR fellowship (n. DOT13IEP1U-1). V.L. is supported by a PNRR fellowship Mission 4, Component 1, Action 4.1 "Research Doctorates", CUP: B53C23001770006 by the European Union-NextGenerationEU.

Data availability

MultiOmicsXplorer web application is available at: https://perfettolab.bio.uniroma1.it/MultiOmicsXplorer/. The tutorial is available at: https://perfettolab.bio.uniroma1.it/MultiOmicsXplorer/tutorial.html.The web application is implemented in R using the Shiny framework. The source code is available at: https://github.com/SaccoPerfettoLab/MultiOmicsXplorer/. The original datasets used in this study were generated by the CPTAC and are described in Li et al., 2023 [13]. The data are publicly available through the CPTAC data portal and the Proteomic Data Commons (PDC) (https://pdc.cancer.gov/), and were accessed in this study via the LinkedOmics web portal (https://www.linkedomics.org/) under the "CPTAC pan-cancer" datasets.All the preprocessed CPTAC data are available at https://perfettolab.bio.uniroma1.it/PerfettoLabData/PatientProfiler/CPTAC/ . Supplementary figures are available online.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 July 2025 / Accepted: 20 April 2026
Published online: 25 April 2026

References

1. Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. Cell. 2020;183:1436–e145631. https://doi.org/10.1016/j.cell.2020.10.036.
2. Anurag M, Jaehnig EJ, Krug K, Lei JT, Bergstrom EJ, Kim B-J, et al. Proteogenomic markers of chemotherapy resistance and response in triple-negative breast cancer. Cancer Discov. 2022;12:2586–605. https://doi.org/10.1158/2159-8290.CD-22-0200.
3. Hyeon DY, Nam D, Han Y, Kim DK, Kim G, Kim D, et al. Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes. Nat Cancer. 2022. https://doi.org/10.1038/s43018-022-00479-7.
4. Hernandez-Valladares M, Vaudel M, Selheim F, Berven F, Bruserud Ø. Proteogenomics approaches for studying cancer biology and their potential in the identification of acute myeloid leukemia biomarkers. Expert Rev Proteom. 2017;14:649–63. https://doi.org/10.1080/14789450.2017.1352474.
5. Akhoundova D, Rubin MA. Clinical application of advanced multi-omics tumor profiling: shaping precision oncology of the future. Cancer Cell. 2022;40:920–38. https://doi.org/10.1016/j.ccell.2022.08.011.
6. Thangudu RR, Holck M, Singhal D, Pilozzi A, Edwards N, Rudnick PA, et al. NCI's proteomic data commons: a cloud-based proteomics repository empowering comprehensive cancer analysis through cross-referencing with genomic and imaging data. Cancer Res Commun. 2024;4:2480–8. https://doi.org/10.1158/2767-9764.CRC-24-0243.
7. Dugourd A, Kuppe C, Sciacovelli M, Gjerga E, Gabor A, Emdal KB, et al. Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. Mol Syst Biol. 2021;17:e9730. https://doi.org/10.15252/msb.20209730.
8. Bradley G, Barrett SJ. CausalR: extracting mechanistic sense from genome scale data. Bioinforma Oxf Engl. 2017;33:3670–2. https://doi.org/10.1093/bioinformatics/btx425.
9. Nilsson A, Peters JM, Meimetis N, Bryson B, Lauffenburger DA. Artificial neural networks enable genome-scale simulations of intracellular signaling. Nat Commun. 2022;13:3069. https://doi.org/10.1038/s41467-022-30684-y.
10. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14:e8124. https://doi.org/10.15252/msb.20178124.
11. Venafra V, Sacco F, Perfetto L. SignalingProfiler 2.0 a network-based approach to bridge multi-omics data to phenotypic hallmarks. NPJ Syst Biol Appl. 2024;10:95. https://doi.org/10.1038/s41540-024-00417-6.
12. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. 2018;46(D1):D956–63. https://doi.org/10.1093/nar/gkx1090.
13. Li Y, Dou Y, Da Veiga Leprevost F, Geffen Y, Calinawan AP, Aguet F, et al. Proteogenomic data and resources for pan-cancer analysis. Cancer Cell. 2023;41(8):1397–406. https://doi.org/10.1016/j.ccell.2023.06.009.
14. Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, et al. Integrated proteogenomic characterization of clear cell renal cell carcinoma. Cell. 2019;179:964–e98331. https://doi.org/10.1016/j.cell.2019.10.007.
15. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell. 2019;177:1035–e104919. https://doi.org/10.1016/j.cell.2019.03.030.
16. Wang L-B, Karpova A, Gritsenko MA, Kyle JE, Cao S, Li Y, et al. Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell. 2021;39:509–e52820. https://doi.org/10.1016/j.ccell.2021.01.006.
17. Huang C, Chen L, Savage SR, Eguez RV, Dou Y, Li Y, et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell. 2021;39:361–e37916. https://doi.org/10.1016/j.ccell.2020.12.007.
18. Satpathy S, Krug K, Jean Beltran PM, Savage SR, Petralia F, Kumar-Sinha C, et al. A proteogenomic portrait of lung squamous cell carcinoma. Cell. 2021;184:4348–e437140. https://doi.org/10.1016/j.cell.2021.07.016.
19. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell. 2020;182:200–e22535. https://doi.org/10.1016/j.cell.2020.06.013.
20. McDermott JE, Arshad OA, Petyuk VA, Fu Y, Gritsenko MA, Clauss TR, et al. Proteogenomic characterization of ovarian HGSC implicates mitotic kinases, replication stress in observed chromosomal instability. Cell Rep Med. 2020;1:100004. https://doi.org/10.1016/j.xcrm.2020.100004.
21. Cao L, Huang C, Zhou DC, Hu Y, Lih TM, Savage SR, et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell. 2021;184:5031–e505226. https://doi.org/10.1016/j.cell.2021.08.023.
22. Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, et al. Proteogenomic characterization of endometrial carcinoma. Cell. 2020;180:729–e74826. https://doi.org/10.1016/j.cell.2020.01.026.
23. Lombardi V, Di Rocco L, Meo E, Venafra V, Di Nisio E, Perticaroli V, Nicolaeasa ML, Cencioni C, Spallotta F, Negri R, Sacco F, Perfetto L. PatientProfiler: building patient-specific signaling models from proteogenomic data. Mol Syst Biol. 2025. https://doi.org/10.1038/s44320-025-00160-y.
24. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? Int J Methods Psychiatr Res. 2011;20:40–9. https://doi.org/10.1002/mpr.329.

25. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45:1–67. https://doi.org/10.18637/jss.v045.i03.
26. Lo Surdo P, Iannuccelli M, Karis K, Meo E, Omidi P, Tosoni M, Graziosi S, Panni S, Fiorillo MT, Licata L, Sacco F, Gyori BM, Perfetto L. SIGNOR 4.0: the 2025 update with focus on phosphorylation data. Nucleic Acids Res. 2026;54(D1):D682–90. https://doi.org/10.1093/nar/gkaf1237.
27. Dugourd A, Saez-Rodriguez J. Footprint-based functional analysis of multiomic data. Curr Opin Syst Biol. 2019;15:82–90. https://doi.org/10.1016/j.coisb.2019.04.002.
28. Diril MK, Ratnacaram CK, Padmakumar VC, Du T, Wasser M, Coppola V, et al. Cyclin-dependent kinase 1 (Cdk1) is essential for cell division and suppression of DNA re-replication but not for liver regeneration. Proc Natl Acad Sci U S A. 2012;109:3826–31. https://doi.org/10.1073/pnas.1115201109.
29. Squire CJ, Dickson JM, Ivanovic I, Baker EN. Structure and inhibition of the human cell cycle checkpoint kinase, Wee1A kinase: an atypical tyrosine kinase with a key role in CDK1 regulation. Struct Lond Engl 1993. 2005;13:541–50. https://doi.org/10.1016/j.str.2004.12.017.

Publisher's note