Systems biology

xOmicsShiny: an R shiny application for cross-omics data analysis and pathway mapping

Benbo Gao[1,#], Yu H. Sun[1,#,*], Xinmin Zhang[2], Tinchi Lin[1], Wei Li[2], Romi Admanit[1], and Baohong Zhang[1,*]

[1]Research and Development, Biogen Inc., Cambridge, MA, USA.

[2]Data Science, BioInfoRx, Inc., Madison, WI, USA.

[#]Contributed equally to the work.

*To whom correspondence should be addressed.

Abstract

Summary: We developed xOmicsShiny, a feature-rich R Shiny-powered application that enables biologists to fully explore omics datasets across experiments and data types with emphasis on uncovering biological insights at the pathway level. The data merging feature ensures flexible exploration of cross-omics data, such as transcriptomics, proteomics, metabolomics, and lipidomics. The pathway mapping function covers a broad range of databases, including WikiPathways, Reactome, and KEGG pathways. In addition, xOmicsShiny offers several visualization options and analytical tasks in everyday omics data analysis, namely, PCA, Volcano plot, Venn Diagram, Heatmap, WGCNA and advanced clustering analyses. The application employs customizable modules to perform various analysis, generating both interactive plots and publication-ready figures. This dynamic, modular design overcomes the slow loading issue of R Shiny tools and allows it to be readily expanded by the research and developer community.

Availability and implementation: The R Shiny application is publicly available at: http://xOmicsShiny.bxgenomics.com. Researchers can upload their own data to the server or use the pre-loaded demo dataset. The source code, under GPLv3 license is provided at https://github.com/interactivereport/xOmicsShiny for local installation. A full tutorial of the application is available at https://interactivereport.github.io/xOmicsShiny/tutorial/docs/index.html.

Contact: yuhenry.sun@biogen.com or baohong.zhang@biogen.com

Supplementary data: Supplementary data are available at *bioRxiv* online.

1 Introduction

With the growing trend of applying high-throughput technologies in biological research, a large amount of data including RNA-Seq, proteomics, and metabolomics has been generated to tackle complex biological questions. Since each omics type represents only one aspect of the biological process, cross-comparison of different types of omics data provides a more comprehensive understanding of how these changes converge at the pathway level. Indeed, multi-omics approaches have been widely used in many research areas (Hasin, et al., 2017), such as developmental biology (Rabinowitz, et al., 2017; Talman, et al., 2018), cancer (Li, et al., 2023; Yuan, et al., 2022), metabolic processes (Williams, et al., 2022), and complex diseases such as Alzheimer's disease (Iturria-Medina, et al., 2022; Kodam, et al., 2023). Furthermore, the multi-omics strategy bolsters the evidence of robust changes under certain conditions, which also facilitates biomarker discovery for diagnostics and therapeutics (Ghosh, et al., 2016; Ren, et al., 2016; Sun, et al., 2019).

Multi-omics datasets are typically processed and summarized into large tables to store quantitative measurements and differential expression analysis results. To gain further insights into the underlying biological changes at the systems biology level, researchers have developed tools to accommodate this challenge, such as MONGKIE (Jang, et al., 2016), WIlsON (Schultheis, et al., 2019), Argonaut (Brademan, et al., 2020), multiSLIDE (Ghosh, et al., 2021), OmicsAnalyst (Zhou, et al., 2021), MiBiOmics (Zoppi, et al., 2021), PaintOmics 4 (Liu, et al., 2022), etc. However, considering the amount and complexity of modern omics data modalities, these tools only partially addressed the bottleneck of data visualization and interpretation. For instance, flexible integration of cross-omics data remains challenging, and a comprehensive pathway mapping function is in demand.

To fill these gaps, we developed xOmicsShiny, a user-friendly R shiny application for interactive analysis of multi-omics datasets. It is designed to advance data exploration and network/pathway-level interpretation for either single-omics or multi-omics data. In particular, it features integrative analysis by merging any gene/protein-related comparisons (such as RNA-Seq, proteomics) or compound-focused datasets (such as metabolomics, lipidomics). Users can navigate WGCNA or PCSF network analysis results and map the data to multiple pathway databases such as WikiPathways, Reactome, and KEGG. xOmicsShiny is powered by multiple

state-of-the-art R packages and users have the flexibility to upload, store, and analyze their data using our public server or on their local computers.

2 Methods

Overall Architecture. xOmicsShiny is a web application written using the Shiny package (http://shiny.rstudio.com/), in R programming language. R Shiny applications can be easily configured on a local computer or deployed on a public or private server, enhancing user accessibility and data privacy. To ensure efficient loading of multiple datasets, we employed a modular design and created 14 modules (Fig. 1) for diverse analyses. Users only need to load the required modules when using them, which saves memory and time during application loading.

Following the best-practice of bioinformatic analysis, we incorporated packages that have received broad recognition by the research community, including ComplexHeatmap (Gu, 2022; Gu, et al., 2016), fgsea (Korotkevich, et al., 2021), WGCNA (Langfelder and Horvath, 2008), PCSF (Akhmedov, et al., 2017), pathview (Luo and Brouwer, 2013), etc.

Data Merging and Meta p-value Calculation. Merging data is a critical step to multi-omics analysis. xOmicsShiny automatically matches gene symbols and compound IDs and joins the differential expression results derived from multiple comparisons. While keeping the original information after merging, additionally, we provide four methods for meta p-value calculation: fisher, minP, simes, and stouffer. The meta fold change is displayed as the average fold change by default, but users can also provide a customized formula to summarize the fold changes.

Enrichment Analysis and Pathway Visualization. As a key step of pathway analysis and prioritization, both Gene Set Enrichment Analysis (GSEA) and Over-Representation Analysis (ORA) are used in the 'Geneset Enrichment' module to calculate the enrichment statistics. The enrichment step is performed using the fgsea package (Korotkevich et al., 2016), and the visualization is facilitated by the collapsePathways function. We currently support WikiPathways, Reactome, and KEGG databases. A few examples of MetaBase pathways have also been included to demonstrate the versatile integration of commercially available databases into xOmicsShiny. The enrichment analysis is compatible with single-omics data (RNA-Seq, proteomics, or metabolomics data) or merged data containing both gene symbols and compound IDs. xOmicsShiny performs a one-step enrichment analysis using the KEGG database and displays both genes and compounds on the KEGG pathway map. To streamline this step, we created a metabolite database by combining multiple databases and constructing look-up tables across various ID formats. A total of 263,000 unique metabolites were extracted from PubChem, KEGG, HMDB, SMPDB, Metabolite Refmet. etc.

Weighted Gene Co-expression Network Analysis (WGCNA). We used the WGCNA R package to perform network analysis when multiple samples are available. The calculation may be time-consuming so it will be run in the background and results will be saved. Users can easily load and explore the results once they are ready.

Dose Response and Time Course Analysis. The Dose Response module offers multiple built-in models such as the log-logistic model, generalized log-logistic model, and Weibull model. The time course analysis automatically fits the data to several models such as linear, Gauss-probit, log Gauss-probit, as well as exponential models.

3 Features

xOmicsShiny is designed for comprehensive multi-omics data visualization, with an emphasis on understanding the biological mechanisms at the systems biology level. To this end, we developed 14 modules under the 'Settings' tab encompassing five major types of tasks: 1) Data merging, 2) Data exploration, 3) GSEA and pathway analysis, 4) Network analysis, and 5) Pattern and trend analysis. We have provided a detailed tutorial on how to process, upload, and navigate the data on xOmicsShiny. Required inputs should include original expression values, sample metadata, and differential expression statistics such as fold change, p-values, and adjusted p-values. Moreover, we curated public data containing RNA-Seq, proteomics, and metabolomics to serve as a demo dataset in the platform hosted by our public server. A set of supplementary figures featuring key functions of the application can be found in Supplementary data (Supplementary Fig. S1-11).

Data Merging. The 'Merge Data' module is designed to merge multiple same or different types of omics data at the comparison level, rather than at the original data level. This can be achieved through a two-step procedure. First, users need to load multiple datasets they intend to merge. Then, under the 'Merge Data' module, users can select the desired comparisons and merge them together. The merged data will be displayed as a new dataset, containing meta fold change and p-values. Merging the same data type from different comparison groups highlights consistent changes across conditions, while cross-omics data integration facilitates pathway-level integrative analysis.

Data Exploration. General data exploration consists of five modules: Quality Control (QC) module, Differentially Expressed Genes (DEGs) module, Heatmap module, Expression module, and Venn Diagram module. We leverage and improve several functions in the single-omics visualization platform, Quickomics (Gao, et al., 2021). The QC module generates various plots including principal component analysis (PCA) plot, sample-sample distance plot, dendrograms, etc. The DEGs module features volcano plot and scatter plot for visualization. For merged data, users can display side-by-side volcano plots or a merged volcano plot. Heatmap has been widely used in presenting high-throughput data, and the Heatmap module enables both static and interactive heatmap generation. Following that, the Expression module contains functions to display individual gene/compound values in various ways, such as boxplot, violin plot, etc. Lastly, users can examine overlapping genes using the Venn Diagram module. These modules are highly customizable, offering plenty of controls to adjust the output figures.

GSEA and Pathway Analysis. The single 'Geneset Enrichment' module runs GSEA or ORA analysis on a variety of pathway and gene ontology databases, and maps multi-omics data to the pathway diagrams. For merged datasets, KEGG pathway maps can plot fold changes of multiple comparisons simultaneously. The enrichment analysis is performed on-the-fly as the users adjust fold change and p-value cutoffs to select genes or compounds. All enrichment statistics and gene lists can be easily saved for further investigation.

Network Analysis. xOmicsShiny offers three types of network analysis, computed by the WGCNA module, the Correlation Network module, and the PCSF module. WGCNA has been widely used to identify co-expression regulatory networks and hub genes for mechanistic discovery. The WGCNA module displays the dendrogram of hierarchical clustering, which allows users to adjust parameters for tree cutting. Following that, corresponding gene clusters will be shown for further investigation. The Correlation Network module, powered by the visNetwork and networkD3 R packages, provides interactive correlation network visualization on genes or compounds based on user-defined cutoffs. Moreover, we incorporated the Prize-collecting Steiner Forest (PCSF) method to further highlight sub-networks and functional units (Akhmedov et al., 2017).

Pattern and Trend Analysis. Identifying patterns and trends has been critical to certain tasks, including analyzing dose-response treatment data or time-course experiment outputs. To achieve this, we developed four

modules: Pattern module, Curve Fitting Dose Response module, Curve Fitting Time Course module, and Monotonic Trend module. The Pattern module identifies clusters on a set of user-defined genes using soft clustering, k-means, or partition around medoid (PAM) strategies. The optimal number of clusters can be determined by methods including silhouette coefficients and Within-Cluster Sum of Squares (WSS) values. The next two modules, as their names suggest, perform dose response curve fitting and time course fitting analysis. The Monotonic Trend module is useful for identifying consistent trends in the data, especially when the changes are not linear.

4 Conclusions

In summary, xOmicsShiny provides a user-friendly solution to bridge the gap between high-dimensional, multi-omics data and biological insights, with high scalability and extensive features compared with other R Shiny applications (Gao, et al., 2021; Munk, et al., 2024; Tien et al., 2024; Zhao and Wang, 2022; Zoppi, et al., 2021) (Supplementary table S1, S2). Users can follow the detailed tutorial to compile their own data tables, which can be uploaded, stored, and analyzed by the Shiny application. xOmicsShiny contains 14 modules to perform various analyses, from general quality control and exploratory analysis to complex tasks and pathway-level visualization. This is ensured by many databases pre-loaded in the application, such as KEGG database, gene ontology database, Reactome pathway database, and WikiPathways, etc. Moreover, it generates high-resolution, publication-quality figures that biologists can use directly. xOmicsShiny is released as an open-source Shiny application, which facilitates the bioinformatics community to further enhance its functionalities.

Conflict of Interest: B.G., Y.H.S., T.L., R.A. and B.Z. are employees of Biogen and hold stocks from the company. X.Z. and W.L. are employees of BioInfoRx, Inc.

References

Akhmedov, M., *et al.* PCSF: An R-package for network-based interpretation of high-throughput data. *PLoS Comput Biol* 2017;13(7):e1005694.

Brademan, D.R., *et al.* Argonaut: A Web Platform for Collaborative Multi-omic Data Visualization and Exploration. *Patterns (N Y)* 2020;1(7).

Gao, B., *et al.* Quickomics: exploring omics data in an intuitive, interactive and informative manner. *Bioinformatics* 2021;37(20):3670-3672.

Ghosh, N., *et al.* Transcriptomics, proteomics and metabolomics driven biomarker discovery in COPD: an update. *Expert Rev Mol Diagn* 2016;16(8):897-913.

Ghosh, S., Datta, A. and Choi, H. multiSLIDE is a web server for exploring connected elements of biological pathways in multi-omics data. *Nat Commun* 2021;12(1):2279.

Gu, Z. Complex heatmap visualization. *iMeta* 2022;1(3).

Gu, Z., Eils, R. and Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32(18):2847-2849.

Hasin, Y., Seldin, M. and Lusis, A. Multi-omics approaches to disease. *Genome Biol* 2017;18(1):83.

Iturria-Medina, Y., *et al.* Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of Alzheimer's disease progression and heterogeneity. *Sci Adv* 2022;8(46):eabo6764.

Jang, Y., *et al.* MONGKIE: an integrated tool for network analysis and visualization for multi-omics data. *Biol Direct* 2016;11(1):10.

Kodam, P., *et al.* Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. *Sci Rep* 2023;13(1):3695.

Korotkevich, G., *et al.* Fast gene set enrichment analysis. *bioRxiv* 2021:060012.

Langfelder, P. and Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.

Li, X., *et al.* Integrated transcriptomics, proteomics and metabolomics-based analysis uncover TAM2-associated glycolysis and pyruvate metabolic remodeling in pancreatic cancer. *Front Immunol* 2023;14:1170223.

Liu, T., *et al.* PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases. *Nucleic Acids Res* 2022;50(W1):W551-W559.

Luo, W. and Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 2013;29(14):1830-1831.

Munk, Katharina, et al. "Holomics-a user-friendly R shiny application for multi-omics data integration and analysis." *BMC bioinformatics* 2024;25(1):93.

Rabinowitz, J.S., *et al.* Transcriptomic, proteomic, and metabolomic landscape of positional memory in the caudal fin of zebrafish. *Proc Natl Acad Sci U S A* 2017;114(5):E717-E726.

Ren, S., *et al.* Integration of Metabolomics and Transcriptomics Reveals Major Metabolic Pathways and Potential Biomarker Involved in Prostate Cancer. *Mol Cell Proteomics* 2016;15(1):154-163.

Schultheis, H., *et al.* WIlsON: Web-based Interactive Omics VisualizatioN. *Bioinformatics* 2019;35(6):1055-1057.

Sun, Q., *et al.* Integration of metabolomic and transcriptomic profiles to identify biomarkers in serum of lung cancer. *J Cell Biochem* 2019;120(7):11981-11989.

Talman, V., *et al.* Molecular Atlas of Postnatal Mouse Heart Development. *J Am Heart Assoc* 2018;7(20):e010378.

Tien, NTN, et al. "EasyPubPlot: a shiny web application for rapid omics data exploration and visualization." *bioRxiv* 2024;2024-11.

Williams, E.G., *et al.* Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Syst* 2022;13(1):43-57 e46.

Yuan, Q., *et al.* Integration of transcriptomics, proteomics, and metabolomics data to reveal HER2-associated metabolic heterogeneity in gastric cancer with response to immunotherapy and neoadjuvant chemotherapy. *Front Immunol* 2022;13:951137.

Zhao, T., and Wang, Z., GraphBio: A shiny web app to easily perform popular visualization analysis for omics data. *Fron Genetics*, 2022;13:957317.

Zhou, G., Ewald, J. and Xia, J. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res* 2021;49(W1):W476-W482.

Zoppi, J., *et al.* MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinformatics* 2021;22(1):6.

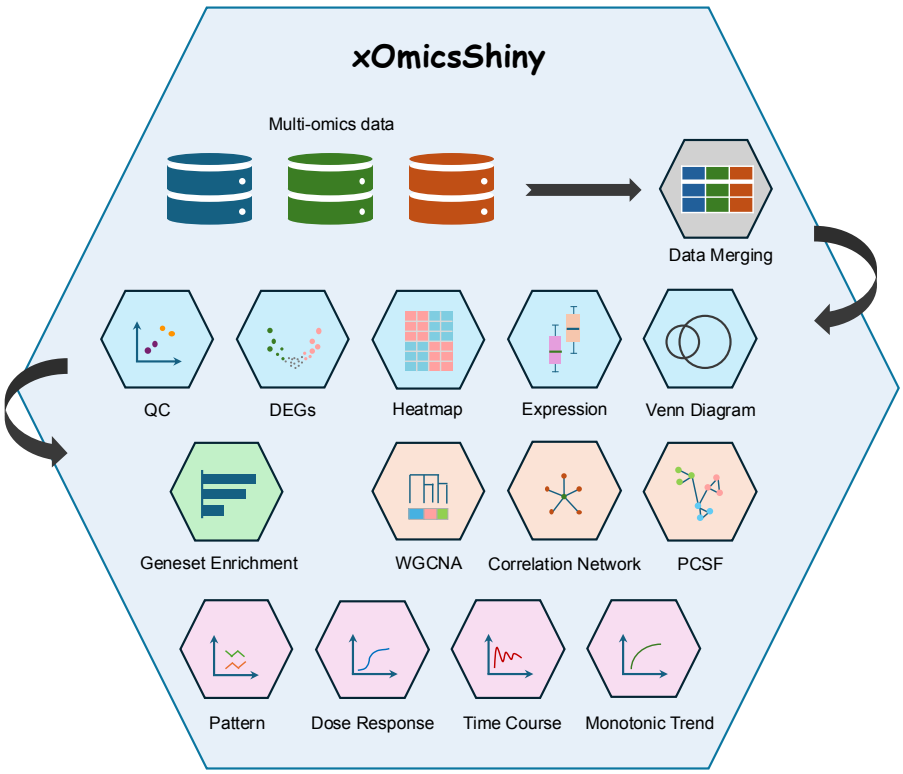
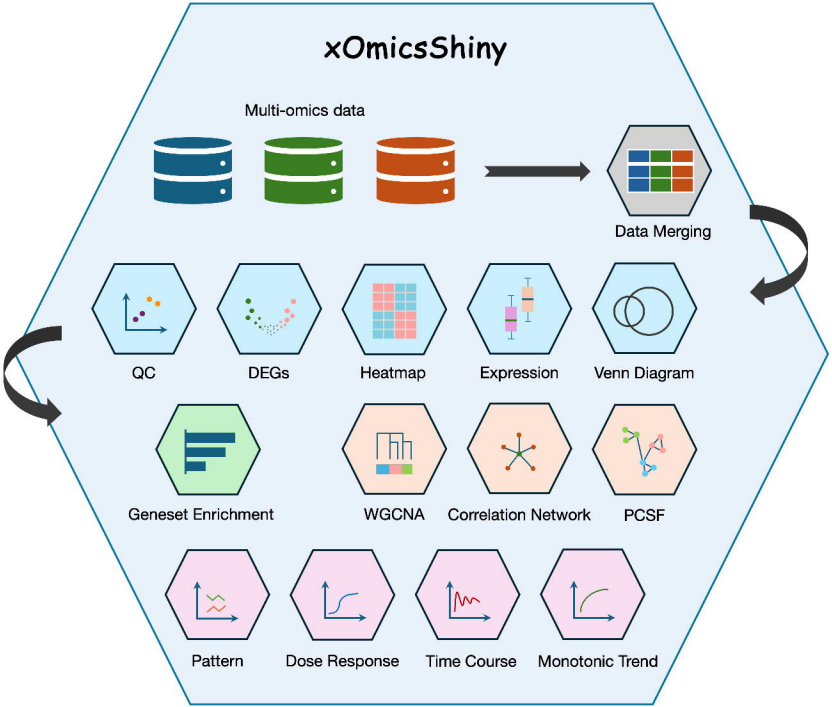


Fig 1. Overview of xOmicsShiny modules. The Shiny application handles multi-omics data including RNA-Seq, proteomics, and metabolomics data. The **Data Merging** (Grey) module integrates multi-omics data for downstream analysis. **Data Exploration** modules (Blue) contain QC, DEGs, Heatmap, Expression, and Venn Diagram modules. **Geneset Enrichment** (Green) module performs GSEA and pathway analysis. **Network Analysis** (Bisque) modules include WGCNA, Correlation Network, and PCSF modules. **Pattern and Trend Analysis** (Orchid) modules consist of Pattern, Dose Response, Time Course, and Monotonic Trend modules.


xOmicsShiny
Multi-omics data
Data Merging
QC
DEGs
Heatmap
Expression
Venn Diagram
Geneset Enrichment
WGCNA
Correlation Network
PCSF
Pattern
Dose Response
Time Course
Monotonic Trend