GraphBio: A shiny web app to easily perform popular visualization analysis for omics data

Tianxin Zhao¹* and Zelin Wang²*

¹Department of Pediatric Urology, Guangzhou Women and Children’s Medical Center, Guangzhou Medical University, Guangdong Provincial Clinical Research Center for Child Health, Guangzhou, Guangdong, China, ²Department of Bioinformatics, Shuzhi Biotech, LLC, Guangzhou, Guangdong, China

Background: Massive amounts of omics data are produced and usually require sophisticated visualization analysis. These analyses often require programming skills, which are difficult for experimental biologists. Thus, more user-friendly tools are urgently needed.

Methods and Results: Herein, we present GraphBio, a shiny web app to easily perform visualization analysis for omics data. GraphBio provides 15 popular visualization analysis methods, including heatmap, volcano plots, MA plots, network plots, dot plots, chord plots, pie plots, four quadrant diagrams, Venn diagrams, cumulative distribution curves, principal component analysis (PCA), survival analysis, receiver operating characteristic (ROC) analysis, correlation analysis, and text cluster analysis. It enables experimental biologists without programming skills to easily perform popular visualization analysis and get publication-ready figures.

Conclusion: GraphBio, as an online web application, is freely available at http://www.graphbio1.com/en/ (English version) and http://www.graphbio1.com/ (Chinese version). The source code of GraphBio is available at https://github.com/databio2022/GraphBio.

KEYWORDS
omics, heatmap, volcano plots, visualization, GraphBio

Introduction

With the advance of high-throughput techniques (Goodwin et al., 2016), more and more researchers have started to depict molecular profiling in a systematic manner (Cancer Genome Atlas, 2012; Yan et al., 2015; Cancer Genome Atlas Research Network. Electronic address and Cancer Genome Atlas Research, 2017; Wang et al., 2018; Zhao

Abbreviations: FDR, false discovery rate; GO, Gene Ontology; ROC, receiver operating characteristic; PCA, principal component analysis.
FIGURE 1
Flow diagram of the retrospective protocol for case inclusion and exclusion and data collection. Some representative visualization results from GraphBio. (A) Gene expression heatmap. Rows represent genes, and columns represent samples. Yellow represents upregulation, and blue represents downregulation. (B) The volcano plot shows significantly differentially expressed genes. The x-axis represents log2-transformed fold changes, and the y-axis represents −log10 (FDR). False discovery rate, FDR. Red represents upregulation, blue represents downregulation, and yellow represents genes that are not statistically significant. Some genes of interest are marked in purple. (C) MA plot shows significantly differentially expressed genes. The x-axis represents mean expression values of genes, and the y-axis represents log2-transformed fold changes. Red represents (Continued)
et al., 2022). Massive amounts of omics data are produced and usually require sophisticated visualization analysis. For example, gene expression studies frequently use heatmaps, volcano plots, and MA plots to characterize expression changes from thousands of genes (Conesa et al., 2016). Moreover, principal component analysis (PCA) and correlation analysis are widely used to estimate similarity or dissimilarity between samples or groups. Although these methods are popular in omics research, they were usually published as R packages, such as ggplot2 (Wickham, 2016), heatmap, survminer, pROC (Robin et al., 2011), FactoMineR (Lê et al., 2008), factoextra, and GOplot (Walter et al., 2015). As an online web application, it is freely available at http://www.graphbio1.com/en/(English version) and http://www.graphbio1.com/(Chinese version) without login requirements. It can be accessed through any web browser, such as Google Chrome and Microsoft Edge. The source code of GraphBio is available at https://github.com/databio2022/GraphBio.

Results and discussion

Overview of GraphBio

GraphBio provides 15 visualization analysis modules, including heatmap, volcano plots, MA plots, network plots, MA plots, chord plots, pie plots, four quadrant diagrams, Venn diagrams, cumulative distribution curves, PCA, survival analysis, ROC analysis, correlation analysis, and text cluster analysis. Some representative visualization results are shown in Figure 1. GraphBio supports four input file formats: csv, txt, xls, and xlsx. Notably, csv files cannot be encoded in UTF-8 and txt files must be tab-separated. All plots can be downloaded in a PDF, PNG, JPEG, or TIFF format with customizable size and resolution. A “run example” button was added to each module for users to learn the corresponding function features quickly without data preparation steps. Users can easily try changing default parameter values to observe the changes in the example figures. The example data are partly shown when users click “view example file” button to facilitate users to prepare their own data files. In addition, we have also added a “Help Center” section as a user manual in GraphBio.

Example 1. Heatmap for gene expression profiles.

Heatmap is a data matrix visualizing values in the cells using a color gradient, and it has been frequently used in omics data analysis. In GraphBio, the “Heatmap” module requires a gene expression matrix file as input. Our example data include 20 genes and 10 samples. A well-prepared csv file was uploaded, and then an idea plot was automatically generated (Figure 1A). We can easily adjust the plot by changing colors,
clustering methods, gene, and sample name presentation as needed. Notably, the module provides six popular color presets, which are commonly presented in many papers, and a colorblind-friendly color was selected as default.

Example 2. Volcano plots for differential expression analysis.

Volcano plots depict the relationship between significance and fold changes of differentially expressed genes, and genes presented in the upper-left and upper-right corners are generally interesting to biologists. The “Volcano plot” module of GraphBio requires an input file with four columns, including geneID, log₂ (fold change), significance (p or padj values), and label. The “label” column represents the genes to be highlighted on the figure. The example result is clearly shown in Figure 1B. The numbers of upregulated and downregulated genes were summarized. We can also customize the colors, fold changes, significance threshold, point size, and label size.

Example 3. Four quadrant diagrams for differential expression analysis between two omics data sets.

Four quadrant diagrams are generally used to analyze two omics data sets, such as RNA-seq and m6A-seq. We used differentially expressed genes and peaks from RNA-seq and m6A-seq data as a demonstrated example. The input file included six columns, including geneID, log₂FoldChanges (RNA-seq), significance (p or padj values, RNA-seq), log₂FoldChanges (m6A-seq), significance (p or padj values, m6A-seq), and label. The “label” column represents genes to be highlighted on the figure. The resulting figure is shown in Figure 1H. Four groups of genes are highlighted in different colors, and corresponding gene numbers are also summarized. We can also adjust the significance threshold, fold changes, point size, label size, and colors.

Conclusion

In this article, we introduce GraphBio, an online web application for omics data visualization. It includes 15 popular visualization analysis methods, such as heatmaps, volcano plots, and MA plots. Experimental biologists can easily perform online analysis and get publication-ready plots via accessing the website http://www.graphbio1.com/en/(English version) or http://www.graphbio1.com/(Chinese version) using any web browsers like Google Chrome and Microsoft Edge. In the future, we will continue integrating more popular visualization analysis methods into GraphBio and provide more support to the research community.

Data availability statement

The original contributions presented in the study are publicly available. These data can be found at: https://github.com/databio2022/GraphBio.

Author contributions

ZW led the software development. TZ tested the software and wrote the initial manuscript. TZ and ZW modified the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (no. 2022A1515012576), the Research Fund for Clinical Doctor of Guangzhou Women and Children’s Medical Center, and the Research Fund from Shuzhi Biotech, LLC, Guangzhou. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Conflict of interest

Author ZW was employed by the company Shuzhi Biotech, LLC.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70. doi:10.1038/nature11412

Cancer Genome Atlas Research Network. Electronic address and Cancer Genome Atlas Research (2017). Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 169, 1327–1341.e23. e1323. doi:10.1016/j.cell.2017.05.046

Chen, T., Liu, Y. X., and Huang, L. (2022). ImageGP: An easy-to-use data visualization web server for scientific researchers. iMeta 1, e5. doi:10.1002/imt2.5
Reyes, A. L. P., Silva, T. C., Coetzee, S. G., Plummer, J. T., Davis, B. D., Chen, S., et al. (2019). GENAVi: A shiny web application for gene expression normalization, analysis and visualization. BMC Genomics 20, 745. doi:10.1186/s12864-019-6073-7

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 12, 77. doi:10.1186/1471-2105-12-77

Walter, W., Sánchez-Cabo, F., and Ricote, M. J. B. (2015). GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics 31, 2912–2914. doi:10.1093/bioinformatics/btv300

Wang, Z. L., Li, B., Luo, Y. X., Lin, Q., Liu, S. R., Zhang, X. Q., et al. (2018). Comprehensive genomic characterization of RNA-binding proteins across human cancers. Cell Rep. 22, 296–298. doi:10.1016/j.celrep.2017.12.035

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag.

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., et al. (2015). Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell. 28, 529–540. doi:10.1016/j.ccell.2015.09.006

Zhao, T., Tang, X., Li, D., Zhao, J., Zhou, R., Shu, F., et al. (2022). Prenatal exposure to environmentally relevant levels of PBDE-99 leads to testicular dysgenesis with steroidogenesis disorders. J. Hazard. Mat. 424, 127547. doi:10.1016/j.jhazmat.2021.127547