ELSEVIER

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

BioVisReport: A Markdown-based lightweight website builder for reproducible and interactive visualization of results from peer-reviewed publications

Jingcheng Yang [a,b,1], Yaqing Liu [a,1], Jun Shang [a,1], Yechao Huang [a], Ying Yu [a], Zhihui Li [a], Leming Shi [a], Zihan Ran [c,d,⁎]

[a] State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences and Shanghai Cancer Center, Fudan University, 2005 Songhu Road, Shanghai 200438, China
[b] Greater Bay Area Institute of Precision Medicine, 115 Jiaoxi Road, Guangzhou 511458, China
[c] Department of Research, Shanghai University of Medicine & Health Sciences Affiliated Zhoupu Hospital, 1500 Zhouyuan Road, Shanghai 201318, China
[d] Inspection and Quarantine Department, The College of Medical Technology, Shanghai University of Medicine & Health Sciences, 279 Zhouzhu Road, Shanghai 201318, China

ARTICLE INFO

Article history:
Received 18 March 2022
Received in revised form 5 June 2022
Accepted 5 June 2022
Available online 8 June 2022

Keywords:
Reproducibility
Website builder
Interactive visualization
Peer-reviewed publications

ABSTRACT

Interactive visualization is an effective way to promote the reproducibility of results presented in biomedical publications and to facilitate additional exploration of the reported data. However, there is a lack of convenient tools that balance reproducibility with ease of use. To address this problem, we develop BioVisReport, a lightweight solution for the rapid generation of an interactive website based on a user-defined Markdown file, which acts as a text markup language without requiring users to master complex syntax and allows them to preview the results in real-time. Interactive websites generated by the tool can help readers conveniently reproduce research findings and perform further in-depth analyses beyond those reported in the original peer-reviewed publications. Currently, BioVisReport offers 17 basic types of plots for visualizing published data. In addition, the extensibility of BioVisReport supports flexible integration of user-developed Python plugins with multiple programming languages. BioVisReport is freely available at https://biovis.report/.

© 2022 The Author(s). 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 license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Reproducibility of results in biomedical publications by a third party is essential but challenging [1]. The disclosure of raw data and analysis pipeline is strongly advocated to facilitate the reproducibility of results by peers [2–4]. However, it is unavoidable that there is still a large amount of research data with low accessibility, which directly limits outside researchers to repeat the analysis of publications [5,6]. Even in the presence of data accessibility, irrational or undocumented manipulations in the analysis process, such as different software versions, parameters, threshold conditions, etc., likewise lead to irreproducibility dilemmas [7,8].

In this context, the idea of using interactive visualisation to drive reproducibility was proposed [9]. This approach has been mainly applied to aggregated, normalized or segmented data, also known as level 3 data, such as variant calling files and quantitative expression profiles [10]. This type of data is widely used for downstream analysis and has the closest connection to the conclusions of a publication. Interactive visualization ties these data to the code for downstream analysis so that readers can analyze the data directly by point-and-click means. It allows the reader to quickly reproduce results in the publication and also perform in-depth analysis based on the interactive figures [11]. Interactive visualization differentiates itself from the traditional approach that uses a limited number of static figures to highlight the most important findings in a publication, making it possible to bridge the gap between the conclusions and the underlying data [12,13].

Common implementations of visualizing research results interactively can be divided into two categories. One is to provide a computational notebook containing interactive visual widgets, e.g., Jupyter Notebooks with ipywidgets [14], which essentially provides readers with code and data for re-analysis. However, improper coding by the authors may lead to a high degree of irre-

⁎ Corresponding author at: Department of Research, Shanghai University of Medicine & Health Sciences Affiliated Zhoupu Hospital, 1500 Zhouyuan Road, Shanghai 201318, China.

E-mail address: ranzh@sumhs.edu.cn (Z. Ran).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.06.009
2001-0370/© 2022 The Author(s). 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 license (http://creativecommons.org/licenses/by/4.0/).

producibility [15,16]. Another more effective way is to provide a stand-alone online website that can be easily accessed through a URL. However, this approach not only requires a high level of programming skills and web development capabilities for the authors but is also time-consuming and labor-intensive.

Here, we developed BioVisReport, a lightweight website builder for interactive visualization reports, to facilitate the implementation of the FAIR (Findability, Accessibility, Interoperability, and Reusability) principle in biomedical publications [17]. Users can generate a website with this tool by writing a Markdown file, a text markup language with the feature of writing what you get and previewing the results in real time [18]. Users can format the text content with some simple markups, such as adding different numbers of "#" at the beginning of the lines to set them as different levels of headers. Researchers can attach the website address to their publication, allowing readers to use it to reproduce the original results and dig further into the data.

2. Materials & methods

BioVisReport invokes specified plugins and biomedical data based on a user-defined Markdown file, and then generates an interactive website. This work is achieved through two components of BioVisReport, i.e., the website generator and the plugin system (Fig. 1).

2.1. Website generator

Website generator is built using Python and additional libraries including Jinja2 [19], Markdown [18], Tornado [20], Mkdocs [21], and livereload [22]. The generation of an interactive website requires the user to prepare data, set the website format and specify plot plugins. In BioVisReport, all the above tasks can be defined in a Markdown file. A syntax recognizer was developed to parse the plugin invocation syntax in the Markdown file and then execute these commands. Each plugin may receive one or more input files. The files can be located on a local disk or remotely, and can be in a tabular data file format (CSV or TSV), Rdata, or other formats. All data files are first downloaded or cached and then loaded using the built-in Python or R libraries. A Markdown converter receives the elements generated by the syntax recognizer and, at the same time, converts other commands in the Markdown file into the corresponding HTML statements to achieve the layout effect expected by the user.

2.2. Plugin system

The plugin system contains a variety of mutually independent plot plugins that can be invoked by the website generator. Our strategy is to develop separate plugins for each type of graph to achieve compatibility among multiple programming languages and a high level of extensibility. Each plugin has been encapsulated as a Python package and can be installed separately. BioVisReport recognizes Python packages through an API interface. This design strategy allows users with programming skills to develop customized plugins that are not yet supported in BioVisReport. All currently available plugins were developed using Python, JavaScript, R and additional libraries including Plotly [23], Dplyr [24], Shiny [25], ggplot2 [26], WebDataRocks [27], etc.

In addition, BioVisReport is designed to accommodate non-expert users by supporting both development and production modes, improving usability through faster feedback. In development mode, the instance follows the principle of immediate feedback, and it will generate the website and embed a Markdown editor into the website. Through the Markdown editor, the user can modify the Markdown file and trigger the instance to respond to modifications in real-time.

3. Results

3.1. Overview of BioVisReport

BioVisReport has been implemented as a Conda package (biovis-report) in development mode for easy installation and as a Docker image (biovis-report-viewer) in production mode for fast deployment and migration. The installation, usage manual and application examples are available at https://biovis.report/.

BioVisReport has integrated 17 basic and commonly used plugins (Table 1). In addition to the basic data table, pivot tables that allow users to filter and combine calculations to produce timely

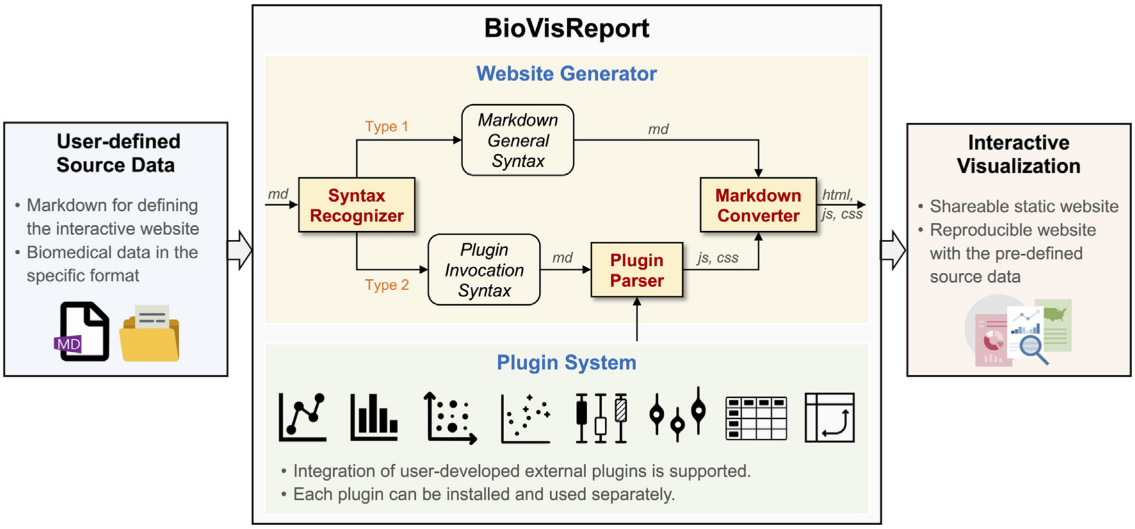

BioVisReport
Website Generator
User-defined Source Data
• Markdown for defining the interactive website
• Biomedical data in the specific format
MD
md
Syntax Recognizer
Type 1
Markdown General Syntax
md
Markdown Converter
html, js, css
Type 2
Plugin Invocation Syntax
md
Plugin Parser
js, css
Plugin System
• Integration of user-developed external plugins is supported.
• Each plugin can be installed and used separately.
Interactive Visualization
• Shareable static website
• Reproducible website with the pre-defined source data

Fig. 1. The architecture of BioVisReport. The Website Generator can recognize the syntax of the Markdown file and divide the commands into two categories: (1) for the Markdown general syntax, it will be converted into elements of the website, i.e., *html*, *js* and *css*; (2) for the commands invocating plugins, a special parser is designed for generating *js* and *css* elements, which will be passed to the Markdown converter to participate in generating the final website. In addition, the Plugin System provides plugins capable of making interactive plots.

Table 1
Summary of interactive plot plugins currently available within BioVisReport.

No.	Interactive Plugin	Usage
1	Data table based on DataTables	@datatable-js (dataUrl)
2	Pivot table based on WebDataRocks and Highcharts	@pivot-table-js (dataUrl, enableLocal)
3	Pie chart based on Echarts	@pie-chart-js (dataUrl, group, subgroup, value, title, radius, chartName, legendOrient, legendPosition, selectedMode)
4	Table based on Tabulator	@tabulator-js (dataUrl)
5	Bar plot from a Shiny app	@barplot-r (dataFile, dataType, title, xAxis, xTitle, yAxis, yTitle, colorAttr, shapeAttr, xLog10, enableSE, showpanel, subtitle, text, queryURL)
6	Box plot from a Shiny app	@boxplot-r (dataFile, dataType, title, xAxis, xTitle, xAngle, yAxis, yTitle, colorAttr)
7	Correlation plot from a Shiny app	@corrplot-r (dataFile, dataType, corrVars, title, xTitle, xAngle, yTitle, corrMethod, corrType, hcOrder, hcMethod, showLab, showPanel)
8	Density plot from a Shiny app	@density-plot-r (dataFile, dataType, title, xAxis, xTitle, colorAttr, subtitle, text, fillEnable)
9	Group-box plot from a Shiny app	@grouped-boxplot-r (dataFile, dataType, title, xAxis, xTitle, yAxis, yTitle, colorAttr, labelAttr, legendTitle, subtitle, text)
10	Heatmap from a Shiny app	@heatmap-d3-r (dataFile, dataType, rowv, colv, distfun, hclustfun, showpanel, scale, labRow, labCol, colNameLst)
11	Line plot from a Shiny app	@lineplot-r (dataFile, dataType, title, xAxis, xTitle, yAxis, yTitle, colorAttr, shapeAttr, xLog10, enableSE, showpanel, subtitle, text, queryURL)
12	Rocket plot from a Shiny app	@rocket-plot-r (dataFile, dataType, title, subtitle, xAxis, xTitle, yAxis, yTitle, xAngle, labelAttr, method, pointAlpha, pointSize, text)
13	Scatter plot from a Shiny app	@scatter-plot-r (dataFile, dataType, xAxis, yAxis, sizeAttr, nameAttr, colorAttr, labelAttr, showpanel)
14	Stack bar plot from a Shiny app	@stack-barplot-r (dataFile, dataType, title, xAxis, xTitle, yAxis, yTitle, xAngle, labelAttr, subtitle, text, barPos, smartColor)
15	Upset plot from a Shiny app	@upset-r (dataFile, dataType, title, showEmptyInterSec, showBarNumbers, setSort, nIntersects, assignmentType, subtitle, text, showpanel)
16	Violin plot from a Shiny app	@violin-plot-r (dataFile, dataType, title, xAxis, xTitle, yAxis, yTitle, xAngle, colorAttr, subtitle, text)
17	MultiQC	@multiqc-py (analysisDir)

statistical charts are provided. Besides, these plugins can be used to draw pie chart, bar plot, box plot, correlation plot, density plot, group-box plot, heatmap, line plot, rocket plot, scatter plot, stack bar plot, upset plot, and violin plot. Finally, MultiQC [28] reports can be integrated, which means that users can quickly develop more comprehensive and richer reports based on the power of MultiQC, thus greatly extending the scope of applications.

3.2. Usage

During the development stage, the design and content of the reports often need to be constantly adjusted. In this case, we provide biovis-report, a Conda package that contains all the dependent plugins and supports live-reloading (Fig. 2A). Users can immediately generate an interactive report based on Markdown files and simple terminal commands, with changes to Markdown reacting to the report in real time. Report generation with biovis-report consists the following steps, including preparing data, designing the Markdown templates, activating Conda environment to run BioVisReport and then obtain the HTML file.

In production mode, biovis-report-viewer is suitable for formal releasing and sharing after the final report has been generated (Fig. 2B). Here, we recommend two approaches to apply the interactive reports generated by BioVisReport to publications. On the one hand, authors can deploy the locally generated reports on a server, i.e., transform a temporary HTML into a URL that can be stably accessed. In particular, for purely static websites entirely based on JavaScript plugins, server deployment is even not required, only the HTML file is stored on a publicly accessible cloud drive. Then authors can attach the website address to the publication. On the

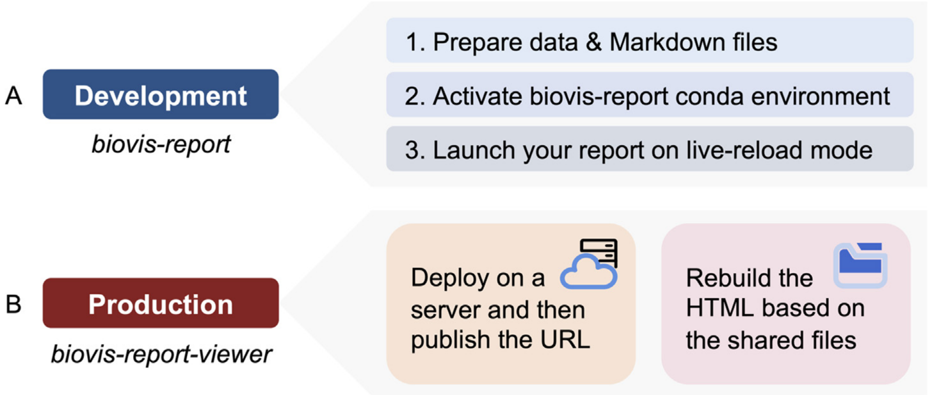
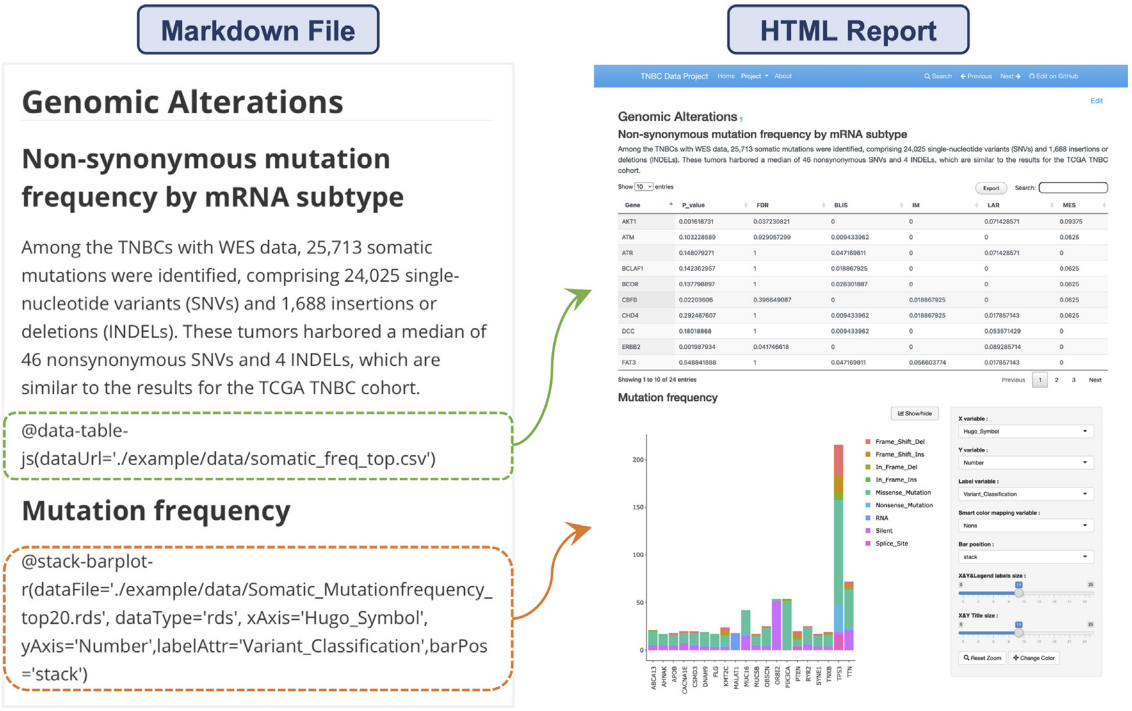

A
Development
biovis-report
1. Prepare data & Markdown files
2. Activate biovis-report conda environment
3. Launch your report on live-reload mode
B
Production
biovis-report-viewer
Deploy on a server and then publish the URL
Rebuild the HTML based on the shared files

Fig. 2. Development and production modes of BioVisReport. (A) In development mode, users prepare the data and Markdown files, and then develop the report in a live-reload state of biovis-report. (B) In production mode, the interactive reports can be published and deployed to the server, or can be reproduced with the original data using biovis-report-viewer.



Markdown File

Genomic Alterations

Non-synonymous mutation frequency by mRNA subtype

Among the TNBCs with WES data, 25,713 somatic mutations were identified, comprising 24,025 single-nucleotide variants (SNVs) and 1,688 insertions or deletions (INDELs). These tumors harbored a median of 46 nonsynonymous SNVs and 4 INDELs, which are similar to the results for the TCGA TNBC cohort.

```
@data-table-js(dataUrl='./example/data/somatic_freq_top.csv')
```

Mutation frequency

```
@stack-barplot-r(dataFile='./example/data/Somatic_Mutationfrequency_top20.rds', dataType='rds', xAxis='Hugo_Symbol', yAxis='Number',labelAttr='Variant_Classification',barPos='stack')
```

HTML Report

Genomic Alterations

Non-synonymous mutation frequency by mRNA subtype

Among the TNBCs with WES data, 25,713 somatic mutations were identified, comprising 24,025 single-nucleotide variants (SNVs) and 1,688 insertions or deletions (INDELs). These tumors harbored a median of 46 nonsynonymous SNVs and 4 INDELs, which are similar to the results for the TCGA TNBC cohort.

Show 10 entries

Gene	P_value	FDR	BLIS	IM	LAR	MES
AKT1	0.001618731	0.037230821	0	0	0.071428571	0.09375
ATM	0.103228589	0.929057299	0.009433962	0	0	0.0625
ATR	0.148079271	1	0.047169811	0	0.071428571	0
BCLAF1	0.142362957	1	0.018867925	0	0	0.0625
BCOR	0.137798897	1	0.028301887	0	0	0.0625
CBFB	0.02203606	0.396649087	0	0.018867925	0	0.0625
CHD4	0.292467607	1	0.009433962	0.018867925	0.017857143	0.0625
DCC	0.18018868	1	0.009433962	0	0.053571429	0
ERBB2	0.001987934	0.041746618	0	0	0.089285714	0
FAT3	0.548841888	1	0.047169811	0.056603774	0.017857143	0

Showing 1 to 10 of 24 entries

Mutation frequency

X variable : Hugo_Symbol

Y variable : Number

Label variable : Variant_Classification

Smart color mapping variable : None

Bar position : stack

X&Y&Legend labels size : 11

X&Y Title size : 11

Fig. 3. The "Genomic Alternations" interface shows the variants through data table and bar chart by mRNA subtype, variant classification and gene symbol. The commands to invoke plugins in the Markdown file and the final outcome are shown by the arrows.

other hand, authors can attach the Markdown file with the corresponding data to their publications and inform the readers to install the docker image biovis-report-viewer. The interactive website can be reproduced on the readers' end with just one line of command.

3.3. Example of interactive visualization report

To better demonstrate the capabilities of BioVisReport, we provide an example based on the genomic and transcriptomic profiling of the Chinese triple-negative breast cancer (TNBC) cohort [29]. The details are described below and the report can be viewed through the Video S1.

This report contains a total of four interfaces, i.e., home page, genomic alterations, gene expression, metadata and clinical information. Among them, the home page contains the abstract of the Chinese TNBC report. It does not invoke any interactive plugins, but is converted directly from the Markdown file. The theme of the report is set via command line parameters and the graphical abstract in the center of the page is stored in the working directory in advance. Other interfaces invoked interactive plugins, and the Markdown information and the corresponding results are as follows.

3.3.1. Genomic Alternations

The "Genomic Alternations" tab summarizes somatic SNPs and INDELs in Chinese TNBC with WES data at the beginning (Fig. 3). Then, the data table summarizes the non-synonymous variant frequency by mRNA subtype, along with a *p*-value and an FDR value. The bar chart displays the gene variant frequency stacked by variant classification. Users can change variables (x or y variable, smart color mapping variable, bar position, x and y legend labels or font size of the title, etc.) with the dropdowns and scrollbars in the sidebar panel.

3.3.2. Gene expression

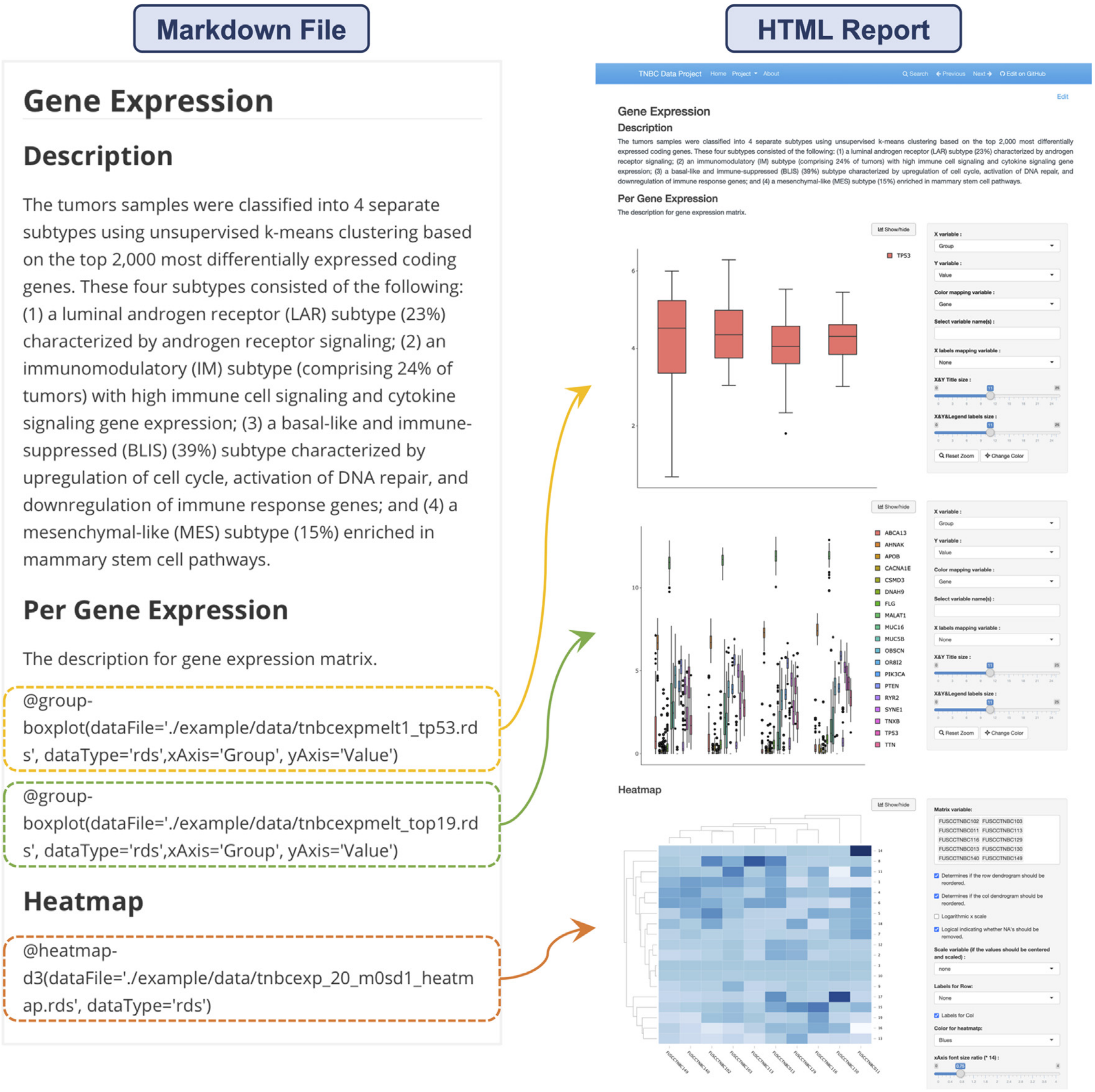
The "Gene Expression" tab integrates heatmap and box plot (Fig. 4). The heatmap allows users to visualize gene-expression data for specified patients selected through the checkbox. This panel also enables users to determine whether to center and scale the value, reorder the dendrogram in row or column, take the x scale logarithm and remove NA values. The box plot shows distribution of the gene-expression levels and users can choose genes they are interested in by clicking on the legend. And the second box plot displays the mapping ratio of two batches of samples.

3.3.3. Meta data and clinical information

The "Meta Data and Clinical Information" tab shows the data type and the clinical data of 495 patients in this cohort (Fig. 5). The data table, which is used to show the clinical data, allows users to query any variable and the pivot table can be used to analyze and organize the complex data. Finally, a scatter plot shows the relationship across clinical variables. These custom features, such as color, symbol or size of mapping variables, arbitrary threshold lines, and confidence ellipses, are provided for users to mine in detail of the correlation between variables.

4. Conclusions

BioVisReport is a lightweight solution for quickly generating an interactive website for reproducing figures associated with a publication. It currently offers 17 basic plugins that can cover the needs of common scenarios of data visualization. Due to the excellent scalability of BioVisReport's system architecture, it allows



Gene Expression

Description

The tumors samples were classified into 4 separate subtypes using unsupervised k-means clustering based on the top 2,000 most differentially expressed coding genes. These four subtypes consisted of the following: (1) a luminal androgen receptor (LAR) subtype (23%) characterized by androgen receptor signaling; (2) an immunomodulatory (IM) subtype (comprising 24% of tumors) with high immune cell signaling and cytokine signaling gene expression; (3) a basal-like and immune-suppressed (BLIS) (39%) subtype characterized by upregulation of cell cycle, activation of DNA repair, and downregulation of immune response genes; and (4) a mesenchymal-like (MES) subtype (15%) enriched in mammary stem cell pathways.

Per Gene Expression

The description for gene expression matrix.

```
@group-boxplot(dataFile='./example/data/tnbcexpmelt1_tp53.rds', dataType='rds',xAxis='Group', yAxis='Value')
```

```
@group-boxplot(dataFile='./example/data/tnbcexpmelt_top19.rds', dataType='rds',xAxis='Group', yAxis='Value')
```

Heatmap

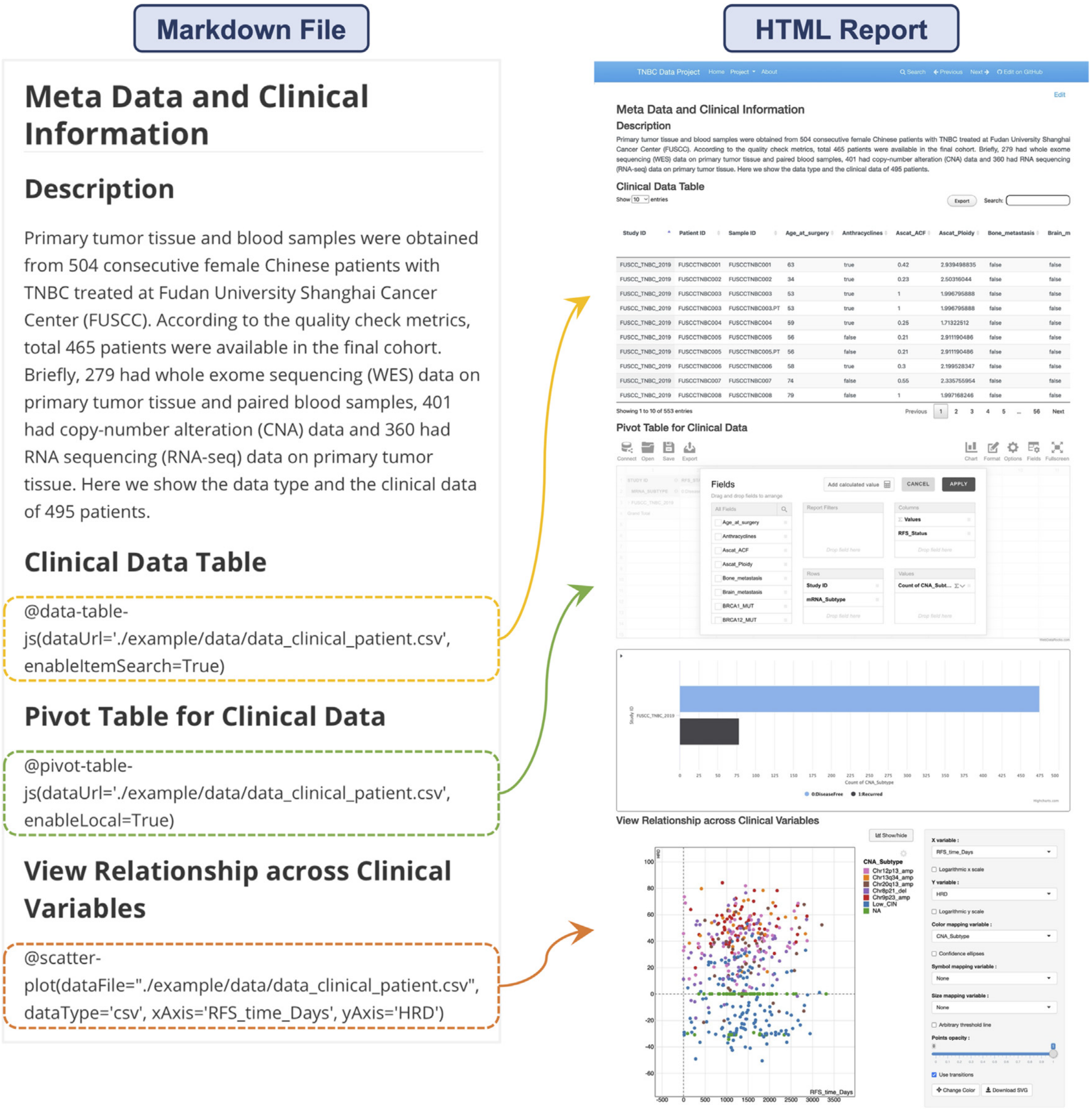
```
@heatmap-d3(dataFile='./example/data/tnbcexp_20_m0sd1_heatmap.rds', dataType='rds')
```

Fig. 4. The "Gene Expression" interface displays data in the form of heatmap and box plot. The commands to invoke plugins in the Markdown file and the final outcome are shown by the arrows.

quick extensions with encapsulating user-defined tools. For example, the integration of a visualization method of a new omic-type data only requires the encapsulation of the existing code as a new plugin. In addition, in order to simplify the development of interactive websites for data exploration and visualization analysis tasks, we have worked on two fronts. On the one hand, we adopted easy-to-use and instantly visible Markdown files in the website design process. On the other hand, during the website modification and optimization process, we provided the development mode so that when users modify data or Markdown files, the ongoing BioVisReport instance can respond to the changes in real-time.

Importantly, BioVisReport has been used by external researchers to build their data portal, further demonstrating the utility of our tool in biomedical peer-reviewed publications [30]. However, more simplified usage, extensive documentation and a wide range of plugins are required for further promotion and adoption. For example, in the current version it is difficult for non-expert users who do not know the Markdown language and do not have an environment, e.g., Anaconda, to carry out its development. We plan to address the issue in a subsequent development of a dashboard that works in a drag-and-drop style. Users can drag the elements and then the Markdown file will be generated automatically. Based on this, we also plan to integrate and provide more visualization plugins.

Overall, the model of applying interactive visualization reporting advocated by BioVisReport, makes the data behind the conclusions less limited by the graphical presentation, thus helping readers to better understand the hypotheses and results conveyed in publications. At the same time, it also helps to identify problems of publication bias and selective reporting, which will improve the accuracy of results generation and interpretation. We believe that BioVisReport will be an effective tool in peer-reviewed publica-



Markdown File

Meta Data and Clinical Information

Description

Primary tumor tissue and blood samples were obtained from 504 consecutive female Chinese patients with TNBC treated at Fudan University Shanghai Cancer Center (FUSCC). According to the quality check metrics, total 465 patients were available in the final cohort. Briefly, 279 had whole exome sequencing (WES) data on primary tumor tissue and paired blood samples, 401 had copy-number alteration (CNA) data and 360 had RNA sequencing (RNA-seq) data on primary tumor tissue. Here we show the data type and the clinical data of 495 patients.

Clinical Data Table

```
@data-table-js(dataUrl='./example/data/data_clinical_patient.csv', enableItemSearch=True)
```

Pivot Table for Clinical Data

```
@pivot-table-js(dataUrl='./example/data/data_clinical_patient.csv', enableLocal=True)
```

View Relationship across Clinical Variables

```
@scatter-plot(dataFile="./example/data/data_clinical_patient.csv", dataType='csv', xAxis='RFS_time_Days', yAxis='HRD')
```

HTML Report

Meta Data and Clinical Information

Description

Primary tumor tissue and blood samples were obtained from 504 consecutive female Chinese patients with TNBC treated at Fudan University Shanghai Cancer Center (FUSCC). According to the quality check metrics, total 465 patients were available in the final cohort. Briefly, 279 had whole exome sequencing (WES) data on primary tumor tissue and paired blood samples, 401 had copy-number alteration (CNA) data and 360 had RNA sequencing (RNA-seq) data on primary tumor tissue. Here we show the data type and the clinical data of 495 patients.

Clinical Data Table

Show 10 entries Export Search:

Study ID	Patient ID	Sample ID	Age_at_surgery	Anthracyclines	Ascat_ACF	Ascat_Ploidy	Bone_metastasis	Brain_m
FUSCC_TNBC_2019	FUSCCTNBC001	FUSCCTNBC001	63	true	0.42	2.939498835	false	false
FUSCC_TNBC_2019	FUSCCTNBC002	FUSCCTNBC002	34	true	0.23	2.50316044	false	false
FUSCC_TNBC_2019	FUSCCTNBC003	FUSCCTNBC003	53	true	1	1.996795888	false	false
FUSCC_TNBC_2019	FUSCCTNBC003	FUSCCTNBC003.PT	53	true	1	1.996795888	false	false
FUSCC_TNBC_2019	FUSCCTNBC004	FUSCCTNBC004	59	true	0.25	1.71322512	false	false
FUSCC_TNBC_2019	FUSCCTNBC005	FUSCCTNBC005	56	false	0.21	2.911190486	false	false
FUSCC_TNBC_2019	FUSCCTNBC005	FUSCCTNBC005.PT	56	false	0.21	2.911190486	false	false
FUSCC_TNBC_2019	FUSCCTNBC006	FUSCCTNBC006	58	true	0.3	2.199528347	false	false
FUSCC_TNBC_2019	FUSCCTNBC007	FUSCCTNBC007	74	false	0.55	2.335755954	false	false
FUSCC_TNBC_2019	FUSCCTNBC008	FUSCCTNBC008	79	false	1	1.997168246	false	false

Showing 1 to 10 of 553 entries Previous 1 2 3 4 5 … 56 Next

Pivot Table for Clinical Data

Fields Add calculated value CANCEL APPLY

Drag and drop fields to arrange

All Fields: □ Age_at_surgery, □ Anthracyclines, □ Ascat_ACF, □ Ascat_Ploidy, □ Bone_metastasis, □ Brain_metastasis, □ BRCA1_MUT, □ BRCA12_MUT

Report Filters: Drop field here

Columns: Σ Values, RFS_Status, Drop field here

Rows: Study ID, mRNA_Subtype, Drop field here

Values: Count of CNA_Subt… Σ, Drop field here

View Relationship across Clinical Variables

X variable : RFS_time_Days

□ Logarithmic x scale

Y variable : HRD

□ Logarithmic y scale

Color mapping variable : CNA_Subtype

□ Confidence ellipses

Symbol mapping variable : None

Size mapping variable : None

□ Arbitrary threshold line

Points opacity : 1

☑ Use transitions

Fig. 5. The "Meta Data and Clinical Information" interface integrates data table, pivot table and scatter plot which allows users to view relationships across clinical variables. The commands to invoke plugins in the Markdown file and the final outcome are shown by the arrows.

tions and even in daily academic communication, thus contributing to the reproducibility of biomedical discoveries.

CRediT authorship contribution statement

Jingcheng Yang: Conceptualization, Methodology, Software, Writing – review & editing, Validation, Resources. **Yaqing Liu:** Conceptualization, Writing – original draft, Resources, Visualization. **Jun Shang:** Conceptualization, Software, Resources. **Yechao Huang:** Software, Resources. **Ying Yu:** Resources, Validation. **Zhihui Li:** Resources, Validation. **Leming Shi:** Conceptualization, Writing – review & editing, Validation. **Zihan Ran:** Project administration, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank members of the laboratory for testing the BioVisReport system and the Genius Medicine Consortium (TGMC) for providing technical support.

Funding

This work was supported in part by the National Natural Science Foundation of China (31720103909 and 32170657), the National Key R&D Project of China (2018YFE0201603), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01) and the 111 Project (B13016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.06.009.

References

[1] Peng RD, Hicks SC. Reproducible Research: A Retrospective. Annu Rev Public Health 2021;42:79–93.

[2] Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature News 2014;505(7485):612.

[3] Alsheikh-Ali AA et al. Public availability of published research data in high-impact journals. PLoS ONE 2011;6(9):e24357.

[4] Hamra GB, Goldstein ND, Harper S. Resource sharing to improve research quality. J Am Heart Assoc 2019;8(15):e012292.

[5] Conesa A, Beck S. Making multi-omics data accessible to researchers. Sci Data 2019;6(1):1–4.

[6] Krassowski M et al. State of the field in multi-omics research: From computational needs to data mining and sharing. Front Genet 2020;11.

[7] Brito JJ et al. Recommendations to enhance rigor and reproducibility in biomedical research. GigaScience 2020;9(6):p. giaa056.

[8] Eisner D. Reproducibility of science: Fraud, impact factors and carelessness. J Mol Cell Cardiol 2018;114:364–8.

[9] Perkel JM. Data visualization tools drive interactivity and reproducibility in online publishing. Nature 2018;554(7690):133–4.

[10] Zhang Z et al. Uniform genomic data analysis in the NCI Genomic Data Commons. Nat Commun 2021;12(1):1–11.

[11] Font Vivanco D et al. Interactive visualization and industrial ecology: applications, challenges, and opportunities. J Ind Ecol 2019;23(3):520–31.

[12] Weissgerber TL et al. Reveal, don't conceal: transforming data visualization to improve transparency. Circulation 2019;140(18):1506–18.

[13] National Academies of Sciences, E. and Medicine, Reproducibility and replicability in science; 2019: National Academies Press.

[14] Kluyver T, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows. vol. 2016; 2016: ELPUB.

[15] Pimentel JF, et al. A large-scale study about quality and reproducibility of jupyter notebooks. In 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR); 2019: IEEE.

[16] Pimentel JF et al. Understanding and improving the quality and reproducibility of Jupyter notebooks. Empirical Software Eng 2021;26(4):1–55.

[17] Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.

[18] Gruber J, Swartz A. Markdown – Daring Fireball; 2014. https://daringfireball.net/linked/2014/01/08/markdown-extension.

[19] Ronacher A. Jinja: a very fast and expressive template engine; 2017. https://github.com/pallets/jinja.

[20] Dory M, Parrish A, Berg B. Introduction to Tornado: Modern Web Applications with Python; 2012: O'Reilly Media, Inc.

[21] Christie T. MkDocs: Project documentation with MarkDown; 2014. https://www.mkdocs.org/.

[22] Yang H, Pfannschmidt R. Livereload server in Python; 2019. https://github.com/lepture/python-livereload.

[23] Johnson A, et al. Plotly: The front end for ML and data science models; 2012. https://plotly.com/.

[24] Wickham H, et al. dplyr: A grammar of data manipulation; 2021. https://CRAN.R-project.org/package=dplyr.

[25] Chang W, et al. shiny: Web Application Framework for R; 2020. https://CRAN.R-project.org/package=shiny.

[26] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

[27] WebDataRocks. A feature-rich JS pivot grid library for creating interactive reports; 2018. https://github.com/WebDataRocks/web-pivot-table.

[28] Ewels P et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32(19):3047–8.

[29] Jiang YZ et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell 2019;35(3):428–440.e5.

[30] Lang GT et al. Characterization of the genomic landscape and actionable mutations in Chinese breast cancers by clinical sequencing. Nat Commun 2020;11(1):5679.