

Vivid Volcano: Empowering Non-Bioinformaticians to Analyze Pre-Processed Omics Data

Tomasz M. Stępkowski, PhD*

* DatViseR – Independent R developer (<https://github.com/DatViseR>)

ORCID: [0000-0002-8454-4421](https://orcid.org/0000-0002-8454-4421)

Summary

Vivid Volcano is an R Shiny web application – an intuitive tool that helps experimental scientists with no bioinformatics background explore and analyze pre-processed omics data. It enables users to perform crucial bioinformatic analyses without the help of specialists. With Vivid Volcano, one can create highly customizable, publication-ready volcano plots and perform comprehensive data exploration, and gene ontology (GO) enrichment analysis. Users can download variously formatted publication-ready plots and neatly formatted tables that adhere to scientific standards. Vivid Volcano empowers both exploratory and explanatory data analysis.

Statement of need

Modern biological research relies heavily on omics technologies (genomics, proteomics, transcriptomics, etc.) that generate large, complex datasets requiring specialized analysis and processing by bioinformaticians – experts with a unique combination of programming, statistics, and data science skills blended with biology domain knowledge(Manzoni et al. 2018). These experts build analysis pipelines which produce pre-processed data conveying metadata on experiments and statistical results for the changes in expression of thousands of proteins, genes, etc. However, to draw valid biological conclusions, this pre-processed data must be explored and explained by subdomain experts who are usually experimentalists who designed the experiments. Those often do not have strong computational skills to effectively explore data and visualize crucial

conclusions. Despite the widespread adoption of omics technologies in biological research, many experimental scientists struggle with data analysis due to technical barriers and communication challenges between computational and experimental disciplines, even with bioinformatician support.

Vivid Volcano was designed with two basic aims:

1. Empowering experimental biologists with a tool that can help them explore and analyze pre-processed omics data on their own
2. Lowering the workload for bioinformaticians who can focus on more statistically and computationally challenging tasks such as integration of complex multiomics data

Vivid Volcano addresses a critical need in the biological research community by empowering wet-lab scientists to independently analyze and interpret preprocessed omics data without requiring programming expertise or extensive bioinformatics support. The application provides an accessible interface for uploading and diagnosing preprocessed omics data, performing customized statistical analyses including gene set enrichment analysis across more than 8,000 GO categories, and generating publication-ready visualizations. Unlike many existing tools, Vivid Volcano maintains a simple and efficient design that does not rely on Bioconductor libraries, making it more accessible and easier to maintain. Vivid Volcano was designed to provide a smooth and intuitive user experience. For more advanced users who would like to validate the under-the-hood processes, the application also implements a process log system that tracks data processing outcomes, facilitating debugging and enabling analysis of app functionality. These features collectively enable experimental biologists to gain deeper insights from their omics data, accelerate research workflows, and produce publication-quality outputs without advanced computational skills.

In summary Vivid Volcano has been designed based on firsthand experience with the challenges faced by experimental biologists working with omics data and aims to provide

comprehensive yet accessible solutions for generating publication-ready outputs from preprocessed omics datasets.

The application is available online at <https://datviser-vivid-volcano.share.connect.posit.cloud/>, making it readily accessible to researchers worldwide.

The source code is available at <https://github.com/DatViseR/Vivid-Volcano>.

Technical description and implementation

Vivid Volcano is built as a modular R Shiny web application leveraging the robust reactive programming model that Shiny provides. The application architecture is characterized by clear separation between the user interface, server logic, and data processing components. The core functionality is implemented in R, with CSS for custom styling and JavaScript for enhanced interactivity. The application utilizes several key R packages, among others: shiny for the web framework (Chang et al. 2024) , shiny.semantic(Stachura et al. 2024) for user interface, DT for interactive data tables(Xie, Cheng, and Tan 2024) , ggplot2(Wickham 2016) and plotly(Sievert 2020) for generating publication-quality visualizations, GT for publication ready tables(Iannone et al. 2024) and shinyjs(Attali 2021) and shinyalert(Attali and Edwards 2024) for improved user experience . For statistical analysis, the application employs basic R operations and custom functions for ontology enrichment that don't rely on Bioconductor dependencies, making the application more accessible and maintainable as a web tool. Data processing is handled through dplyr, tidyr, and other tidyverse packages which provide efficient data manipulation capabilities[Wickham et al. (2023)](Wickham, Vaughan, and Girlich 2024) . The application implements a custom logging system to track all data transformations, ensuring transparency and reproducibility of results. Vivid Volcano's modular design allows for straightforward extension and maintenance, with clearly separated UI modules for data upload, visualization, statistical analysis, and results export. The application is deployed on Posit Connect Cloud (formerly RStudio Cloud), making it accessible via web browser without requiring local installation.

Statistics and Limitations

Vivid Volcano takes an innovative approach to gene set enrichment analysis by prioritizing speed and accessibility without sacrificing essential statistical rigor. Unlike many similar tools, it does not rely on Bioconductor libraries, which makes the application significantly faster, more lightweight, and better suited for web-based deployment and non-technical user. This independence from heavy computational dependencies allows researchers to analyze data quickly in a browser interface, without requiring specialized software installation. Standard Bioconductor packages typically operate on complex data structures and specialized objects that, while powerful for comprehensive analyses, introduce significant computational overhead unsuitable for responsive web applications. Vivid Volcano addresses this limitation by storing crucial Gene Ontology information in highly optimized binary parquet files—a column-oriented data format designed for efficient querying and retrieval. This streamlined approach provides rapid access to approximately 8,000 GO categories and their gene associations without the memory and processing demands of traditional Bioconductor implementations. Vivid Volcano’s gene set enrichment analysis is based on the hypergeometric test, a robust method for determining whether specific gene sets are overrepresented among regulated genes. This test calculates the probability of observing the particular overlap between regulated genes and genes belonging to specific GO categories by chance. By comparing the actual number of regulated genes in a category against what would be expected randomly, the test provides a measure of enrichment. The implementation utilizes four key values: the total genes detected in the experiment, the number of genes in a given GO category, the total regulated genes in the experiment, and the number of regulated genes found in that category. To control false positives, Vivid Volcano implements multiple safeguards: multiple testing correction, customizable significance thresholds, fold enrichment filtering, and gene set size limitations that exclude very small or overly broad categories. The application also transparently addresses the limitations of standard p-value adjustment methods when applied to Gene Ontology data. As noted in the footnotes of the results tables, traditional approaches like

107 Benjamini–Hochberg (BH) were not developed specifically for hierarchical data structures
108 like the GO tree. Thus, biologically meaningful “parent” categories might be missed when
109 BH over-corrects due to many related, statistically significant “child” categories.
110 Conversely, overly broad categories can become significant when BH under-corrects, driven
111 primarily by very highly significant specialized subcategories.

112 Instead of implementing complex solutions that might impede performance and
113 confuse users, Vivid Volcano offers a smooth user experience while clearly communicating
114 methodological considerations. The application provides a pragmatic balance between
115 statistical precision and practical utility, making sophisticated analyses accessible to
116 researchers without extensive bioinformatics expertise.

Figures

Upload your data

Upload a CSV or TSV file

Maximum file size: 25MB

Browse...

stats_DESeq2_small.csv

Upload complete

Download demo data

(separator:tab,decimal:comma)

Header

Separator

Comma

Semicolon

Tab

Decimal Point

Dot

Comma

Upload

Select Data

Select p-value column

pvalue_WT_KAISO_WT_CTRL

Select regulation column - log2(fold)

log2FoldChange_WT_KAISO_WT_CTRL

Select gene symbols column

hgnc_symbol

Columns Checked ✓

Reset Selection

Figure 1. Data upload and data select and curation modules of the Vivid Volcano application

State of data source preview

Show 10

entriesSearch:

Fold	pvalue	qvalue	Sequence	Proteins	Leading razor protein	Gene names	Protein names
0.128764093	0.527601532	0.47528757	AAESLADPTEYENLFPLGK	P35606:A0A7I2V258;H0Y938;H0YAC7	P35606	COPB2	Coatamer subunit beta
-0.025836267	0.097555613	0.888897025	AASADSTTEGTPADGFTVLSTK	O75367	O75367	H2AFY	Core histone macro-H2A.1
-0.058202565	0.168147368	0.787501926	AASDIAMTELPPTHPIR	P62258;K7EM20	P62258	YWHAE	14-3-3 protein epsilon
0.059923351	0.470498342	0.671513196	AATLILEPAGR	Q86TX2;G3V4F2;A0A2R8Y7H3;A0A087X0W7	Q86TX2	ACOT1;ACOT2	Acyl-coenzyme A thioesterase 1
0.188453883	0.819205711	0.31053951	AATPQVAAELEK	O95900	O95900	TRUB2	Probable tRNA pseudouridine synthase 2
-0.225386441	1.750072533	0.151245714	AAVEDSGTTVETIK	Q5VTR2;C9JWJ4;C9J0A5;C9JXC9	Q5VTR2	RNF20	E3 ubiquitin-protein ligase BRE1A
-0.092001975	0.143305155	0.773546527	ACMLTATAHDGDNVISWSR	Q9BQ67;M0QX71	Q9BQ67	GRWD1	Glutamate-rich WD repeat-containing protein 1
-0.149078012	0.196231484	0.680421053	ADIFVDPVLHTACALDIK	Q92896	Q92896	GLG1	Golgi apparatus protein 1
0.133380353	0.320458873	0.590739274	ADLLSLTQPGREEGSPLELER	P08195;F5GZIO;F5GZS6;J3KPF3	P08195	SLC3A2	4F2 cell-surface antigen heavy chain
0.126187533	1.681475426	0.301296089	ADMDDQFTASISETPVDVR	Q06481	Q06481	APLP2	Amyloid-like protein 2

Previous

1

2

3

4

5

...

98

Next

Figure 2. Interactive datatable preview of the analyzed data in the Vivid Volcano application.

6

Analysis Options

Customize GSEA and volcano plot options

Customize p-value adjustment

Benjamini-Hochberg

Significance Threshold

0.05

GSEA analysis controls

☐ I want to run GSEA

Customize volcano plot annotations

☒ Color significantly regulated genes

Up-regulated color

#FF7081

Down-regulated color

#7973FA

☒ Visualize GO Categories

Select from ~8000 unique GO categories

cytosol adaptive immune response

plasma membrane

Color for cytosol

#440154

Color for adaptive immune response

#00008B

Color for plasma membrane

#FFD700

Number of Gene Labels (0-100)

10

☒ Trim Multiplied Gene Names to First Occurrence

☒ Label your chosen genes

Select gene names to label

ATAD3B MTND1P23 TNFRSF18

Color for custom labeled genes

#000000

Customize plot title

Vivid Volcano

Customize X-axis label

Log2 Fold Change (Condition X vs. Condition Y)

[Draw Volcano Plot](#)

[Download Process Log](#)

Figure 3. Volcano plot customisation options in the Vivid Volcano application

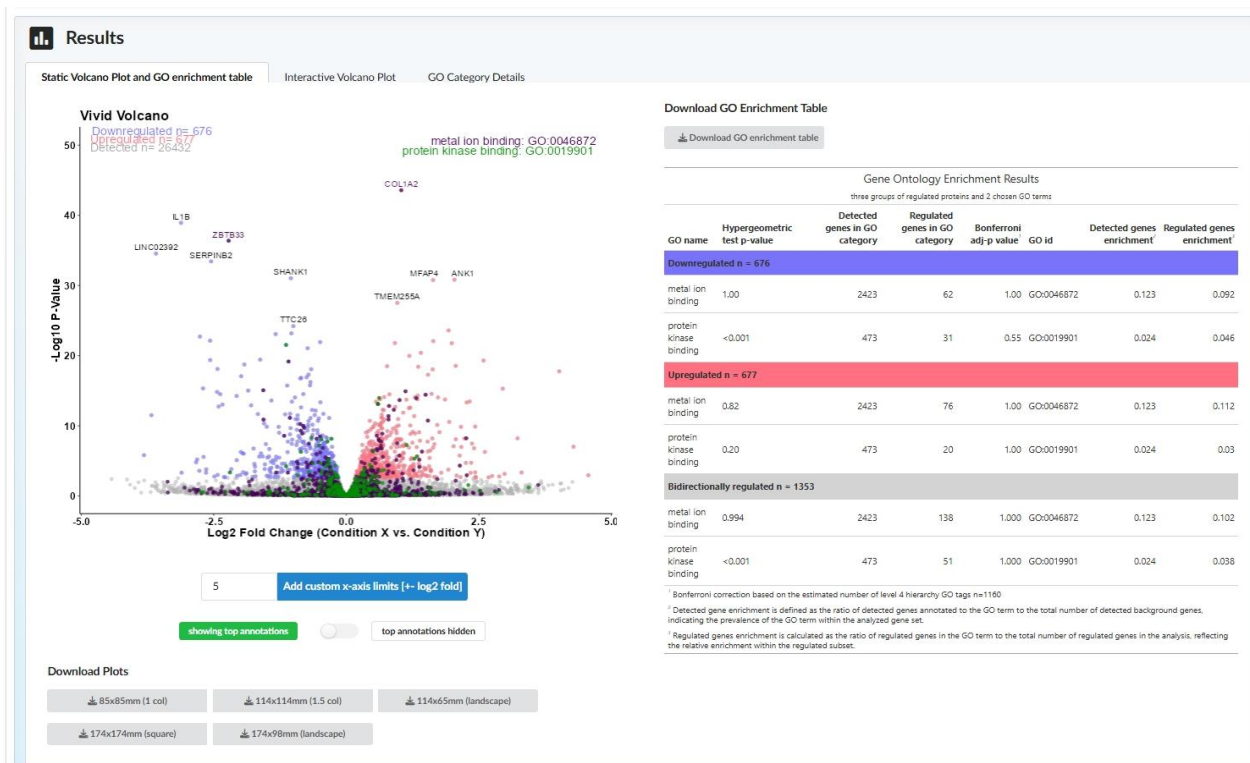
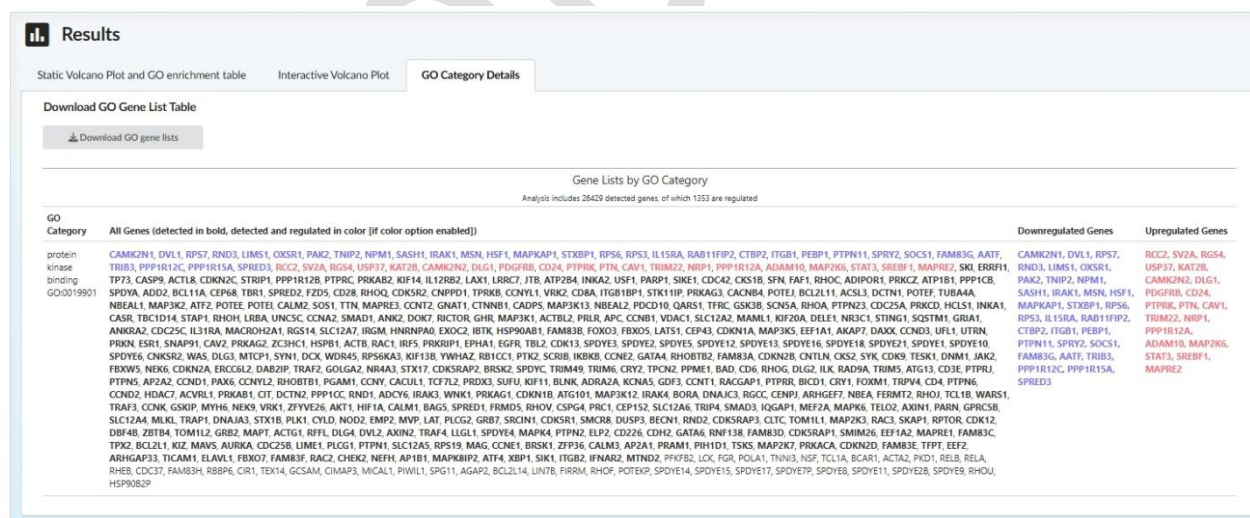


Figure 4. Volcano plot and publication ready tabular results generated with Vivid Volcano application



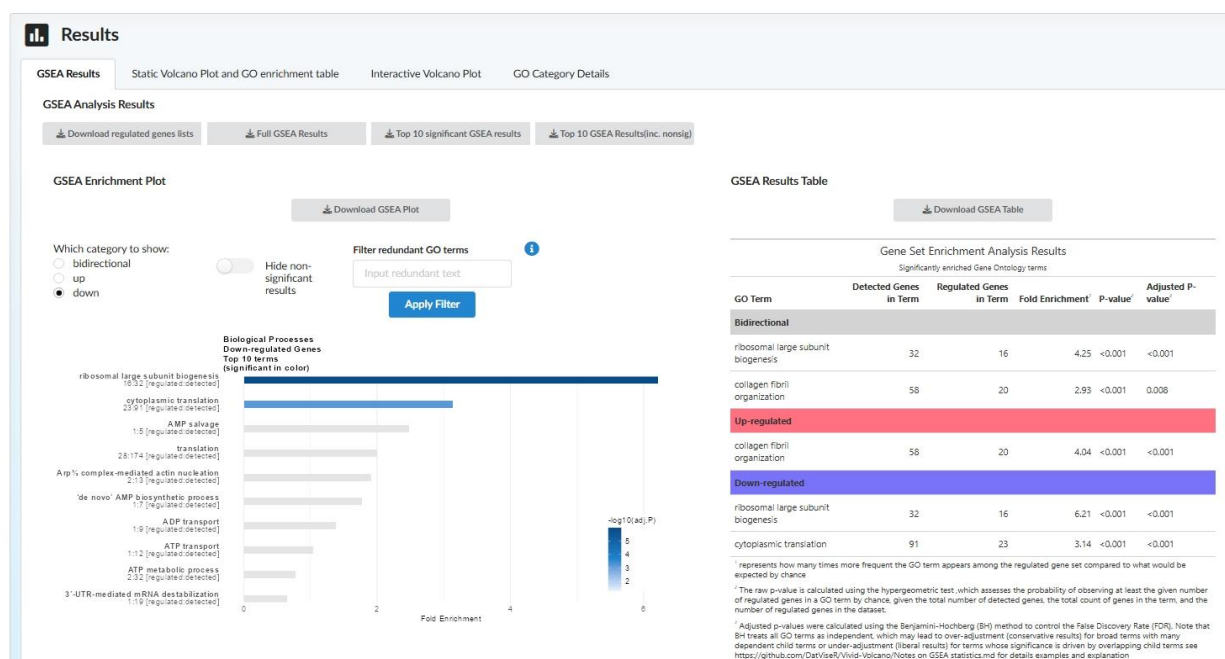


Fig.6. The results of Gene Set Enrichment analysis performed with Vluid Volcano application

Acknowledgements

The author would like to thank all the beta users especially Konrad Kowalski , dr Anna Marusiak and Dr Bartosz Wojtaś for their feedback, suggestions or test data that helped to improve the application.

The author would like to thank the developers of R, RStudio, and the R community for providing the tools and support necessary to create this work.

AI statement

The author used Claude Sonnet 3.7 to copy edit and proofread the manuscript. The author has reviewed generated fragments and corrections to the text to ensure its accuracy and coherence.

References

- Attali, Dean. 2021. “Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds.” <https://deanattali.com/shinyjs/>.
- Attali, Dean, and Tristan Edwards. 2024. “Shinyalert: Easily Create Pretty Popup Messages (Modals) in ‘Shiny’.” <https://github.com/daattali/shinyalert>.
- Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2024. “Shiny: Web Application Framework for r.” <https://CRAN.R-project.org/package=shiny>.
- Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. “Gt: Easily Create Presentation-Ready Display Tables.” <https://gt.rstudio.com>.
- Manzoni, Claudia, Demis A Kia, Jana Vandrovcova, John Hardy, Nicholas W Wood, Patrick A Lewis, and Raffaele Ferrari. 2018. “Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences.” *Briefings in Bioinformatics* 19 (2): 286–302. <https://doi.org/10.1093/bib/bbw114>.
- Sievert, Carson. 2020. “Interactive Web-Based Data Visualization with r, Plotly, and Shiny.” <https://plotly-r.com>.
- Stachura, Filip, Dominik Krzeminski, Krystian Igras, Adam Forys, Paweł Przytuła, Jakub Chojna, Olga Mierzwa-Sulima, Jakub Nowicki, and Tymoteusz Makowski. 2024. “Shiny.semantic: Semantic UI Support for Shiny.” <https://appsilon.github.io/shiny.semantic/>.
- Wickham, Hadley. 2016. “Ggplot2: Elegant Graphics for Data Analysis.” <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. “Dplyr: A Grammar of Data Manipulation.” <https://dplyr.tidyverse.org>.

- 171 Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. “Tidyr: Tidy Messy Data.”
172 <https://tidyr.tidyverse.org>.
- 173 Xie, Yihui, Joe Cheng, and Xianying Tan. 2024. “DT: A Wrapper of the JavaScript Library
174 ‘DataTables’.” <https://github.com/rstudio/DT>.

Preprint