COMMENTARY

iMeta WILEY

OmicStudio: A composable bioinformatics cloud platform with real-time feedback that can generate high-quality graphs for publication

Fengye Lyu | Feiran Han | Changli Ge | Weikang Mao | Li Chen | Huipeng Hu | Guoguo Chen | Qiulei Lang | Chao Fang

Operation Department, LC-Bio Technology Co., Ltd., Hangzhou, China

Correspondence: Fengye Lyu and Chao Fang, LC-Bio Technology Co., Ltd., Hangzhou, China.
Email: fylv@lc-bio.com and cfang@lc-bio.com

In the past decade, an increasing number of cloud-based bioinformatics platforms were spring up, such as Qiita [1], EasyMAP [2], ImageGP [3], MG-RAST [4], gcMeta [5], ETCM [6], Sangerbox [7], antiSMASH [8], EVenn [9], and Majorbio Cloud [10]. These platforms greatly facilitated the biological, medicine and metagenomics research, and provided new vision into big data. Some software deploy various independent tools [11–16], while the others provide complete pipeline [17–21]. The main highlight of these cloud platforms is to convert complicated coding works into easy-used web-based tools [22]. With the assumption that all users already have enough knowledge about bioinformatics before operation, these tools mainly focus on simplifying operation procedure. But the unnoticed truth is that users are also learning while operating, who at the same time expect to get graphs for publication through simple operation.

Here, we display the overall framework as well as each independent module of OmicStudio Cloud Platform (https://www.omicstudio.cn). OmicStudio is a one-stop online analysis platform providing high-throughput omics data analysis, as well as an exploratory platform for bioinformatics research. OmicStudio can obtain results rapidly, generate high-quality graphs for publication, and connect with downstream analysis module automatically. It focuses on speed, quality together with flexibility. Generally, OmicStudio can not only meet the users' demand of ordinary bioinformatics data analysis, statistics, and visualization but also provides them freedom of data mining beyond developer's framework. Additionally, unlimited to developer's aesthetics, users can get more elegant graphs through customizing.

In terms of website framework, OmicStudio has common configuration that all bioinformatics cloud platform may have, such as Cloud Tools, Cloud Analysis, Omics Courses and Study Materials Center, providing entire bioinformatics service for users.

Since its establishment in early 2019, OmicStudio already attracted over 30,000 users, helped whom publish more than 600 research papers (search "OmicStudio" in Google scholar on January 1, 2023), involving various "-omics" fields. An easy-to-use platform which can generate high-quality results is obviously able to make great contribution to scientific research.

OMICSTUDIO OVERVIEW

OmicStudio consists of six modules, named Cloud Tools, Cloud Analysis, Cloud Classroom, Study Materials Center, User's Paper Collection, and User Center. Cloud Tools and Cloud Analysis are the main part of the function for bioinformatics analysis. Cloud Classroom and Study Materials Center help users understand and learn bioinformatics. User's Paper Collection collects the

papers published in cooperation with LCBIO and users can take them as reference. User Center is an integrated management module for user's project information. The core functions of the cloud platform are cloud tools and cloud analytics. The main difference is that cloud tools focus on short, quick analysis, while cloud analytics can support time-consuming analysis and have a project management back office. The two modules are distinguished by the nature of the analysis: the analysis with large amount of data and complex calculation needs to be able to run in the background so that users can leave the web page and do their own thing; simple analytics require instant results and can be personalized immediately. Based on this consideration, the module design can allow users to obtain the most comfortable experience in different application scenarios. Other modules serve these two core modules. Cloud Classroom and Study Materials Center provide the basic knowledge of run bioinformatics analysis in the form of video and text respectively. User Center manages the analysis results, and User's Paper Collection provide new users with reference for writing articles. The innovativeness and character of OmicStudio major function will be clarified later.

In terms of type of omics and application scenarios, Cloud Tools are classified as: General Omics, Single Cell, Enrichment Analysis, Common Analysis, Databases, and so on (Figure 1). The summary of the useful tool in OmicStudio s are showed in Figure 1.

OMICSTUDIO FEATURES

Effect of parameter adjustment with real-time feedback

"Set the parameter and start to analyze," ordinary online bioinformatics tools usually follow this principle. First, users are required to finish the parameter adjustment with unpredictable result, which is with great difficulty. The process of learning that most of us are used to be the "adjust—view the results—adjust again—view the results again" loop. After parameter adjustment and analysis, users need to view the results in backstage. If readjustment is needed, users need to click back to the previous web page, readjust the settings, and view the results again. This is a tedious process with a lot of repetition. Second, for those unfamiliar with bioinformatics, the corresponding relationship between parameter and result is unclear. Consequently, they cannot acquire the desired result through adjusting parameters. Either they are required to spend much time learning the meaning

of parameter, which has nothing to do with their goal, or they must do repetitive work repeatedly. Thirdly, there is no uniform standard for development of bioinformatics website. As a result, users must spend time learning how to use these various online tools. Usually, the function of a bioinformatics tool may be simplified to improve user experience, thus users could get desired results without paying too much time cost. But this simplification will not only limit users' cognition (they don't know how much key information is hidden from their simple operation) but also restrict these online tools to becoming multi-functional platforms, which can provide more personalized bioinformatics tools to meet the demand of more advanced scientific research.

OmicStudio Cloud Platform focuses on real-time feedback. The results of parameter adjustment can be visualized immediately just on the current web page, which greatly reduces users' time cost of debugging. Furthermore, OmicStudio helps users understand the corresponding relationship between parameters and results in an interesting way, turning tedious data analysis into a game. In this game, users can fully understand bioinformatics and how the tools on this platform work. Real-time feedback makes the influence of parameter more comprehensible without knowing its connotation, and then helps data users understand the meaning of parameter through its influence. OmicStudio overcomes the difficulty for users in understanding the esoteric principles of bioinformatics and tools operation, which is a great breakthrough in development of bioinformatics platform.

Providing default parameter, one-click analyze

As mentioned above, the more parameters result in the higher cost of learning. To solve this problem, OmicStudio provides default parameters. In the most convenient way, users only need to upload the documents, wait for the refreshment of the image on the web page (it usually takes just a few seconds), and then download. Despite great number of parameters, each of them has a default value, with which OmicStudio can immediately generate a graph for publication. If users just want to have a glance at their data quality (e.g., using principal component analysis to examine the sample repetition, or using correlation heatmap to check the relevance of data), they can directly download and use, which can be done within 1 min. Default parameters are parameters most frequently used, which are reliable in most cases, and will not be the barrier to choose.

FIGURE 1 Overview tools in OmicStudio. Colors indicate different omics. Pink, General Omics; blue, Single Cell; green, Enrichment Analysis; yellow, Common Analysis; purple, Databases; orange, correlation; bluish violet, microbiome; dark blue, tumor WES. Shape of the icons indicates tool category: circle—Tools, which response immediately; square—Analysis, which runs as a workflow and takes relatively long time.

High-quality and free adjustable analysis results

Ordinary bioinformatics platforms focus only on the analysis procedure. As a result, for publication purposes, the generated graph is usually needed to be redrawn and readjusted. OmicStudio also provides the adjustment module for drawing graphs. The default style of graphs is well designed and adjustable. Users can adjust the themes, color schemes, fonts, titles, and so on. If OmicStudio cannot meet all the needs, it also provides a further solution: downloadable figures in PowerPoint format. The figure generated from OmicStudio can be split into elements in PowerPoint, each of which can be adjusted independently. The function provided by PowerPoint can all be applied to the adjustment of graphs. It has the same function with Adobe Illustrator, while with lower cost of learning.

Associate with downstream analysis module freely

For scientific research, besides the free adjustable parameters, the further demand is to freely choose what to do in the downstream analysis. Although ordinary bioinformatics cloud platforms can provide easy-to-use analysis procedure, they may limit users to freely choosing what to analyze. The analysis modules of OmicStudio are independent while correlated. As shown below, on the page of one kind of cloud tool, there is a series of related cloud tools displayed on the right sidebar, and users can click it to jump to the page of analysis (analysis can be continued based on current results, and there is no necessary to upload any document). The advantage of it is that users can freely choose what to analyze next, as well as decide when to stop. Figure 2 displays the correlation among cloud tools for Tumor WES. Similar series of modules includes Correlation Analysis, Enrichment Analysis, and so on.

High-quality graph for publication

Graphs generated by OmicStudio are famous for its beautiful design. Typical examples are shown in Figure 3, which were all downloaded directly from OmicStudio without any extra adjustment. Graphs from OmicStudio are so well designed that they can still meet the standard of aesthetic even after some customized adjustment. In addition, OmicStudio
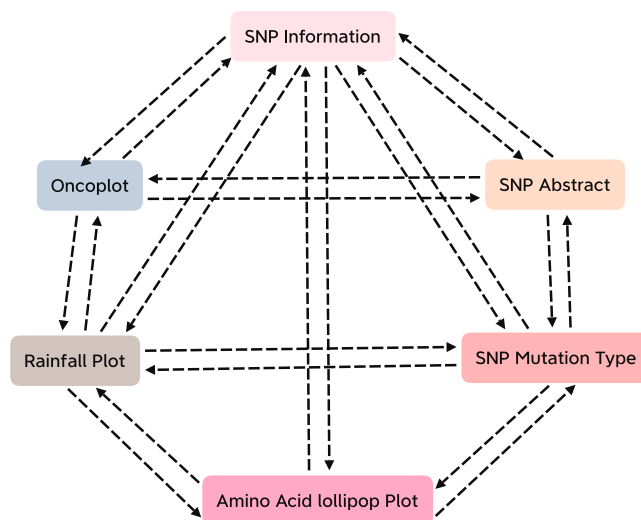

FIGURE 2 Flowchart illustrating the relationship between cloud tools

Cloud Platform supports graphic parameter adjustment, such as fonts, color schemes, themes, and axis adjustment. It's convenient for users to customize the graph according to their demand.

Automatically save the analysis parameter, convenient to backtrack

Ordinary bioinformatics cloud tools save the analysis parameter in background, which is invisible to users. When writing a research paper, it is difficult for users to get the value of parameter from the bioinformatics software.

OmicStudio, on the contrary, provides users parameters along with analysis results. Whether how much time goes by, whether how many versions have been updated, users can always find the parameters in their analysis. Downloadable records of analysis include: reference, analysis method, version of software, analysis parameter, basic information of analysis (when and which tools are used in the platform), citation. This series of information establish a foundation for users' writing.

EXTRA RESOURCES FOR USERS

Cloud Classroom module includes 9 themes and nearly 100 courses. It's convenient to focus on any course based on demand. The themes include bioinformatics, omics science knowledge and literature interpretation, aiming to meet all the users' demand from understanding of biology to the learning of operation.
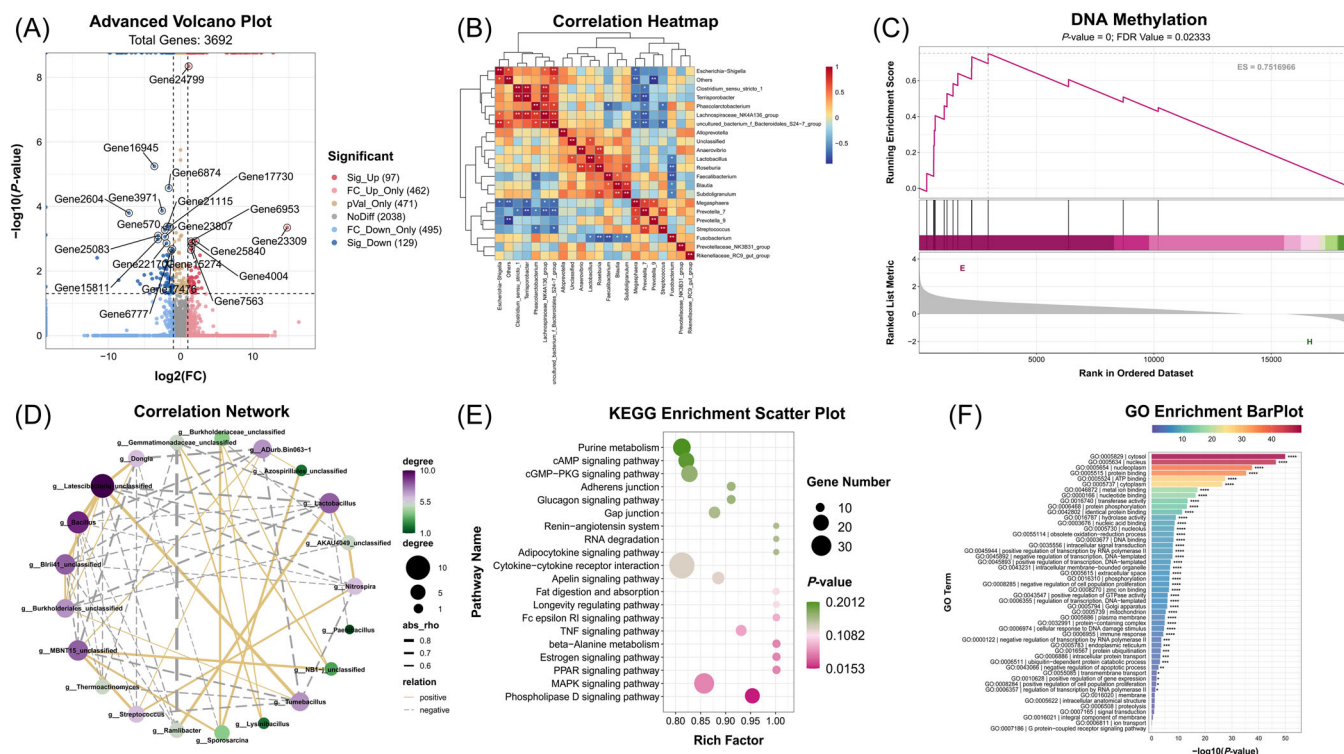
FIGURE 3 Examples of graphs: (A) volcano plot, (B) correlation heatmap, (C) GSEA Enrichment, (D) correlation network, (E) KEGG enrichment scatter plot, and (F) Classic Enrichment.

User's Paper Collection module collects hundreds of research articles published by OmicStudio, providing name of journal, year of publication, impact factor, source of sample, method used and other information. Users can search, preview, and download pdf version of paper. It's convenient for users to acquire knowledge in interested field, and reproduce the analysis result by OmicStudio.

OmicStudio focuses on making scientific research more easily. Since October 2019, the services content of the platform has been upgraded and iterated over 1000 times. There are already more than 30,000 scientific research users. In the past 3 years, 669 research articles cited OmicStudio (Google Scholar, by January 1, 2022).

AUTHOR CONTRIBUTIONS

Fengye Lyu designed the platform and idea. Fengye Lyu and Feiran Han wrote the manuscript. Feiran Han was responsible for editing and revising the manuscript. All authors contributed to the development of OmicStudio Platform.

CONFLICTS OF INTEREST STATEMENT

Qiulei Lang is the Shareholders of LC-Bio Technology Co. The other authors are employees of LC-Bio.

DATA AVAILABILITY STATEMENT

There is no data available.

REFERENCES

1. Gonzalez, Antonio, Jose A. Navas-Molina, Tomasz Kosciolek, Daniel McDonald, Yoshiki Vázquez-Baeza, Gail Ackermann, Jeff DeReus, et al. 2018. "Qiita: Rapid, Web-Enabled Microbiome Meta-Analysis." *Nature Methods* 15: 796–98. https://doi.org/10.1038/s41592-018-0141-9

2. Hung, Yuan-Mao, Tzu-Pin Lu, Mong-Hsun Tsai, Liang-Chuan Lai, and Eric Y. Chuang. 2021. "EasyMAP: A User-Friendly Online Platform for Analyzing 16S Ribosomal DNA Sequencing Data." *New Biotechnology* 63: 37–44. https://doi.org/10.1016/j.nbt.2021.03.001

3. Chen, Tong, Yong-Xin Liu, and Luqi Huang. 2022. "ImageGP: An Easy-to-Use Data Visualization Web Server for Scientific Researchers." *iMeta* 1: e5. https://doi.org/10.1002/imt2.5

4. Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer. 2016. "MG-RAST, A Metagenomics Service for Analysis of Microbial Community Structure and Function." In *Microbial Environmental Genomics (MEG)*, edited by F. Martin and S. Uroz, 207–33. New York: Springer. https://doi.org/10.1007/978-1-4939-3369-3_13

5. Shi, Wenyu, Heyuan Qi, Qinglan Sun, Guomei Fan, Shuangjiang Liu, Jun Wang, Baoli Zhu, et al. 2019. "gcMeta: A Global Catalogue of Metagenomics Platform to Support the Archiving, Standardization and Analysis of Microbiome Data." *Nucleic Acids Research* 47: D637–48. https://doi.org/10.1093/nar/gky1008

6. Xu, Hai-Yu, Yan-Qiong Zhang, Zhen-Ming Liu, Tong Chen, Chuan-Yu Lv, Shi-Huan Tang, Xiao-Bo Zhang, et al. 2018. "ETCM: An Encyclopaedia of Traditional Chinese Medicine." *Nucleic Acids Research* 47: D976–82. https://doi.org/10.1093/nar/gky987

7. Shen, Weitao, Ziguang Song, Xiao Zhong, Mei Huang, Danting Shen, Pingping Gao, Xiaoqian Qian, et al. 2022. "Sangerbox: A Comprehensive, Interaction-Friendly Clinical Bioinformatics Analysis Platform." *iMeta* 1: e36. https://doi.org/10.1002/imt2.36

8. Blin, Kai, Simon Shaw, Katharina Steinke, Rasmus Villebro, Nadine Ziemert, Sang Yup Lee, Marnix H. Medema, and Tilmann Weber. 2019. "antiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline." *Nucleic Acids Research* 47: W81–7. https://doi.org/10.1093/nar/gkz310

9. Chen, Tong, Haiyan Zhang, Yu Liu, Yong-Xin Liu, and Luqi Huang. 2021. "EVenn: Easy to Create Repeatable and Editable Venn Diagrams and Venn Networks Online." *Journal of Genetics and Genomics* 48: 863–66. https://doi.org/10.1016/j.jgg.2021.07.007

10. Ren, Yi, Guo Yu, Caiping Shi, Linmeng Liu, Quan Guo, Chang Han, Dan Zhang, et al. 2022. "Majorbio Cloud: A One-Stop, Comprehensive Bioinformatic Platform for Multiomics Analyses." *iMeta* 1: e12. https://doi.org/10.1002/imt2.12

11. Zhang, Jingying, Yong-Xin Liu, Xiaoxuan Guo, Yuan Qin, Ruben Garrido-Oter, Paul Schulze-Lefert, and Yang Bai. 2021. "High-Throughput Cultivation and Identification of Bacteria From the Plant Root Microbiota." *Nature Protocols* 16: 988–1012. https://doi.org/10.1038/s41596-020-00444-7

12. Yu, Ke, Zhiguang Qiu, Rong Mu, Xuejiao Qiao, Liyu Zhang, Chun-Ang Lian, and Chunfang Deng, et al. 2021. "Recovery of High-Qualitied Genomes from a Deep-Inland Salt Lake Using BASALT." *bioRxiv*. https://doi.org/10.1101/2021.03.05.434042

13. Qian, Xu-Bo, TongChen, Yi-Ping Xu, Lei Chen, Fu-Xiang Sun, Mei-Ping Lu, and Yong-Xin Liu. 2020. "A Guide to Human Microbiome Research: Study Design, Sample Collection, and Bioinformatics Analysis." *Chinese Medical Journal* 133: 1844–55. https://doi.org/10.1097/cm9.0000000000000871

14. Wu, Hua-Jun, Ying-Ke Ma, Tong Chen, Meng Wang, and Xiu-Jie Wang. 2012. "PsRobot: A Web-Based Plant Small RNA Meta-Analysis Toolbox." *Nucleic Acids Research* 40: W22–8. https://doi.org/10.1093/nar/gks554

15. Shi, Wenyu, Guomei Fan, Zhihong Shen, Chuan Hu, Juncai Ma, Yuanchun Zhou, Zhen Meng, et al. 2022. "gcCov: Linked Open Data for Global Coronavirus Studies." *mLife* 1: 92–5. https://doi.org/10.1002/mlf2.12008

16. Liu, Yong-Xin, Yuan Qin, Xiaoxuan Guo, and Yang Bai. 2019. "Methods and Applications for Microbiome Data Analysis." *Hereditas (Beijing)* 41: 845–26. https://doi.org/10.16288/j.yczz.19-222

17. Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2." *Nature Biotechnology* 37: 852–57. https://doi.org/10.1038/s41587-019-0209-9

18. Liu, Yong-Xin, Yuan Qin, Tong Chen, Meiping Lu, Xubo Qian, Xiaoxuan Guo, and Yang Bai. 2021. "A Practical Guide to Amplicon and Metagenomic Analysis of Microbiome Data." *Protein & Cell* 12: 315–30. https://doi.org/10.1007/s13238-020-00724-8

19. Zhou, Zhichao, Patricia Q. Tran, Adam M. Breister, Yang Liu, Kristopher Kieft, Elise S. Cowley, Ulas Karaoz, and Karthik Anantharaman. 2022. "METABOLIC: High-Throughput Profiling of Microbial Genomes for Functional Traits, Metabolism, Biogeochemistry, and Community-Scale Functional Networks." *Microbiome* 10: 33. https://doi.org/10.1186/s40168-021-01213-8

20. Li, Wenjun, Likun Wang, Xiaofang Li, Xin Zheng, Michael F. Cohen, and Yong-Xin Liu. 2022. "Sequence-Based Functional Metagenomics Reveals Novel Natural Diversity of Functioning CopA in Environmental Micro-biomes." *Genomics, Proteomics & Bioinformatics*. https://doi.org/10.1016/j.gpb.2022.08.006

21. de Nies, Laura, Sara Lopes, Susheel Bhanu Busi, Valentina Galata, Anna Heintz-Buschart, Cedric Christian Laczny, Patrick May, and Paul Wilmes. 2021. "PathoFact: A Pipeline for the Prediction of Virulence Factors and Antimicrobial Resistance Genes in Metagenomic Data." *Microbiome* 9: 49. https://doi.org/10.1186/s40168-020-00993-9

22. Koster, Johannes, and Sven Rahmann. 2012. "Snakemake—A Scalable Bioinformatics Workflow Engine." *Bioinformatics* 28: 2520–22. https://doi.org/10.1093/bioinformatics/bts480