



OPEN ACCESS

EDITED BY

Fan Zhang,
University of Colorado Anschutz Medical
Campus, United States

REVIEWED BY

Iain Konigsberg,
University of Colorado System, United States
Yunju Jeong,
Kyung Hee University, Republic of Korea

*CORRESPONDENCE

Ting Wang

✉ tingwang@hospital.cqmu.edu.cn

Zhou Fu

✉ fu_zhou79@126.com

RECEIVED 06 July 2024

ACCEPTED 30 September 2024

PUBLISHED 17 October 2024

CITATION

Peng B, Luo T, Fu X, Zhou Y, Fu Z and Wang T
(2024) Idbview: a database and interactive
platform for respiratory-associated disease.
Front. Immunol. 15:1460422.
doi: 10.3389/fimmu.2024.1460422

COPYRIGHT

© 2024 Peng, Luo, Fu, Zhou, Fu and Wang.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Idbview: a database and interactive platform for respiratory-associated disease

Bingming Peng^{1,2,3,4}, Tingting Luo^{1,2,3,4}, Xingmeng Fu^{1,2,3,4},
Yingzhen Zhou^{1,2,3,4}, Zhou Fu^{1,2,3,4*} and Ting Wang^{1,2,3,4*}

¹Department of Respiratory, Children's Hospital of Chongqing Medical University, Chongqing, China,

²National Clinical Research Center for Child Health and Disorders, Chongqing, China, ³Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China, ⁴Chongqing Key Laboratory of Pediatrics, Children's Hospital of Chongqing Medical University, Chongqing, China

Public databases have become invaluable resources for disease research, particularly in the realm of identifying and validating biomarkers, thus playing a significant role in enhancing our understanding of respiratory diseases. To facilitate this understanding, the development of user-friendly analytical tools and advanced systematic models that leverage the growing omics data and clinical information datasets is essential. Despite the importance of such resources, the research progress related to respiratory diseases is hindered by the absence of a centralized platform housing easily accessible datasets and accompanying visualization tools. In an effort to streamline and standardize information sharing across diverse respiratory research initiatives, we introduce Idbview, a specialized digital database focusing on respiratory conditions, offering interactive visualization functionalities powered by both Vue and R Shiny applications. Idbview brings together clinical data and various omics datasets, serving as a centralized repository, while also providing users with a suite of interactive tools to analyze and visualize data from multiple perspectives. As a comprehensive resource hub, Idbview aims to support the research community in conducting further studies in both clinical and bioinformatics domains, with the website accessible at <https://idbview.com>.

KEYWORDS

database, web application, machine learning (ML), respiratory, Idbview

1 Introduction

Respiratory diseases represent a significant global health challenge, contributing substantially to morbidity and mortality rates. It has been revealed that approximately 545 million individuals are currently affected by respiratory ailments, resulting in 3.9 million deaths annually on a global scale, underscoring the immense burden they place on public health (1). Notably, respiratory diseases account for a substantial portion of hospitalizations in children, directly causing the deaths of 9 million children under the age of 5 each year (2).

During disease research, increasing number of researchers are sharing their experimental data on public databases for validation and further research by the scientific community. These public databases have emerged as invaluable tools for disease investigations, particularly in the realm of identifying and validating biomarkers, thereby saving significant time and resources for subsequent researchers. Nonetheless, a dearth of comprehensive databases focusing on respiratory diseases persists, highlighting the pressing need for their development.

The prevalent database types in use consist of clinical databases and omics databases. The vast majority of clinical databases are not openly accessible, while most omics databases, notably the National Genomics Data Center (3), Gene Expression Omnibus (GEO) (4), and the European Bioinformatics Institute (EBI) (5), are multi-system databases which include multi-organ system diseases, such as cardiovascular system, respiratory system, urinary system, etc. Consequently, there is a scarcity of disease-specific databases devoted to the respiratory system that integrate both clinical and omics data. Furthermore, the disparate data formats and structures across various databases compel researchers to dedicate extensive time and effort to data collection and processing when utilizing multiple databases. Additionally, the lack of extensibility in most databases, absence of compatible data analytics plugins, complex user interfaces, and specific file export formats all contribute to the challenging nature of data querying and analysis processes.

The term “machine learning” encompasses the process of developing predictive models from data or identifying significant patterns within datasets. Various machine learning techniques, including logistic regression, conditional inference trees, random forest (RF), and support vector machine (SVM), have been employed in biology for several decades. The utilization of machine learning in biology has progressively gained importance, becoming a ubiquitous tool across various biological disciplines (6). Nonetheless, these techniques typically necessitate specialized statistical knowledge and programming proficiency, posing a challenge for researchers without a foundational understanding of statistics and programming. Therefore, the development of user-friendly platforms for applying these analytical methods can significantly facilitate their adoption by researchers.

We have developed Idbview (<https://idbview.com>), a database and analysis platform tailored for respiratory diseases. Idbview comprises five main modules: the Clinical Data module, the RNAseq module, the scRNA module, the GraphMed module, and Other. The GraphMed module, a data analysis tool component, was launched on the Hiplot platform (7) (<https://hiplot.com.cn>) and garnered over 20,000 interactions within one year. Following feedback from volunteers and users of the Hiplot platform, enhancements were made to the application’s design and mapping, leading to the integration and development of Idbview version 1. This platform aims to aggregate and standardize data from various sources while offering a user-friendly interactive analysis interface. Moreover, we have seamlessly integrated four machine learning algorithms into the database, enhancing researchers’ ability to leverage these techniques for analyzing extensive omics datasets. Our vision is that Idbview will address the scientific community’s demand for a more accessible,

comprehensive, and in-depth comprehension of the respiratory system, thereby advancing respiratory research.

2 Materials and methods

2.1 Data collection

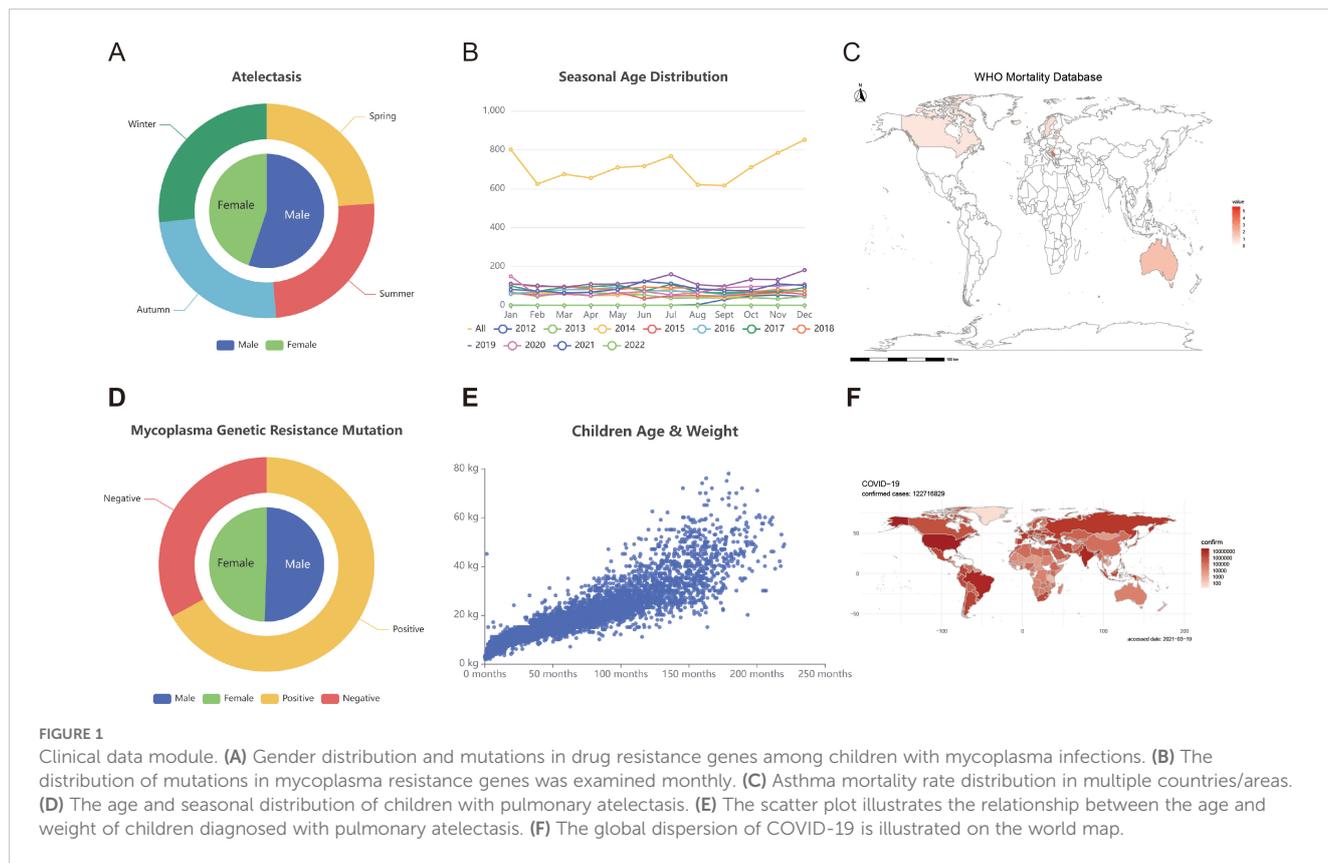
The data were categorized into two types, clinical data and Omics data.

Clinical data: Clinical data were gathered from various sources, including the Children’s Hospital of Chongqing Medical University (CHCMU) in Chongqing, China, World Health Organization (WHO), and GitHub. We collected 8527 cases of childhood atelectasis and 348 cases of mycoplasma-resistant disease in children from CHCMU. This study received ethical approval from the Ethics Committee at CHCMU (2023 Ethics Committee (Research) No. 491). We extracted country-specific mortality data for eight respiratory diseases from WHO, including asthma, Chronic Obstructive Pulmonary Disease (COPD), Covid19, tuberculosis, low respiratory infection, upper respiratory infection, trachea bronchus lung cancers, and other respiratory disease. We obtained mortality data in 127 million cases of Covid-19 and the corresponding visualization code from GitHub (<https://github.com/GuangchuangYu/nCov2019>) (8). To address the variation in formats and content among different databases, the data underwent meticulous processing and standardization procedures. The complete dataset is accessible at <https://idbview.com>. For visualization purposes, we utilized E-charts and R Shiny (Figure 1).

Omics data: We submitted RNA sequence data from 30 samples of 16HBE cell which conducted by the Majorbio Cloud (9). These samples were modelled asthma using house dust mite. The generated data has been deposited for access. Additionally, over 10 thousand respiration-related RNA-seq samples were curated from the GEO, and these datasets were normalized using either the quantile method from the limma package or the variance stabilizing transformation (vst) method from DESeq2 for subsequent analysis. Single-cell RNA (scRNA) data were sourced from the Human Cell Atlas (10) and the Lung Cell Atlas (11) websites. Azimuth (GitHub link: <https://github.com/satijalab/azimuth>) or other user-friendly visualization interfaces were utilized for scRNA data visualization.

2.2 User interface creation

In the construction of the Idbview website, a modern technology stack was employed to ensure optimal efficiency and stability. The frontend development utilized the Vue.js framework along with Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript to create user-friendly interfaces. For backend functions, R Shiny was utilized, and MySQL was chosen for its robust data storage capabilities. Interactive visualizations on the site were created using E-charts in Vue and bs4Dash (12), DESeq2 (13), limma (14), ggplot (15), htmltools (16), and plotly (17) packages in R, among others.



2.3 Technical validation

The database component of Idbview underwent usability testing with the assistance of multiple volunteers, engaging in tasks such as data retrieval, differential analysis, gene screening, and machine learning. For each section, we have added instructions for users to understand the analysis process.

3 Idbview web application

The Idbview interface was created to enable user-friendly access and exploration of the Idbview database. This application can be categorized as 5 modules that can facilitate the investigation of respiratory diseases.

The Clinical Data module contains data on atelectasis and mycoplasma from CHCMU, mortality data of eight respiratory diseases from WHO, and mortality data of Covid-19 from GitHub. The data on atelectasis and mycoplasma are clinical data from cross-sectional visits, including variables such as age, sex, season, and weight. We visualize these data, and users can download these data except for data from CHCMU.

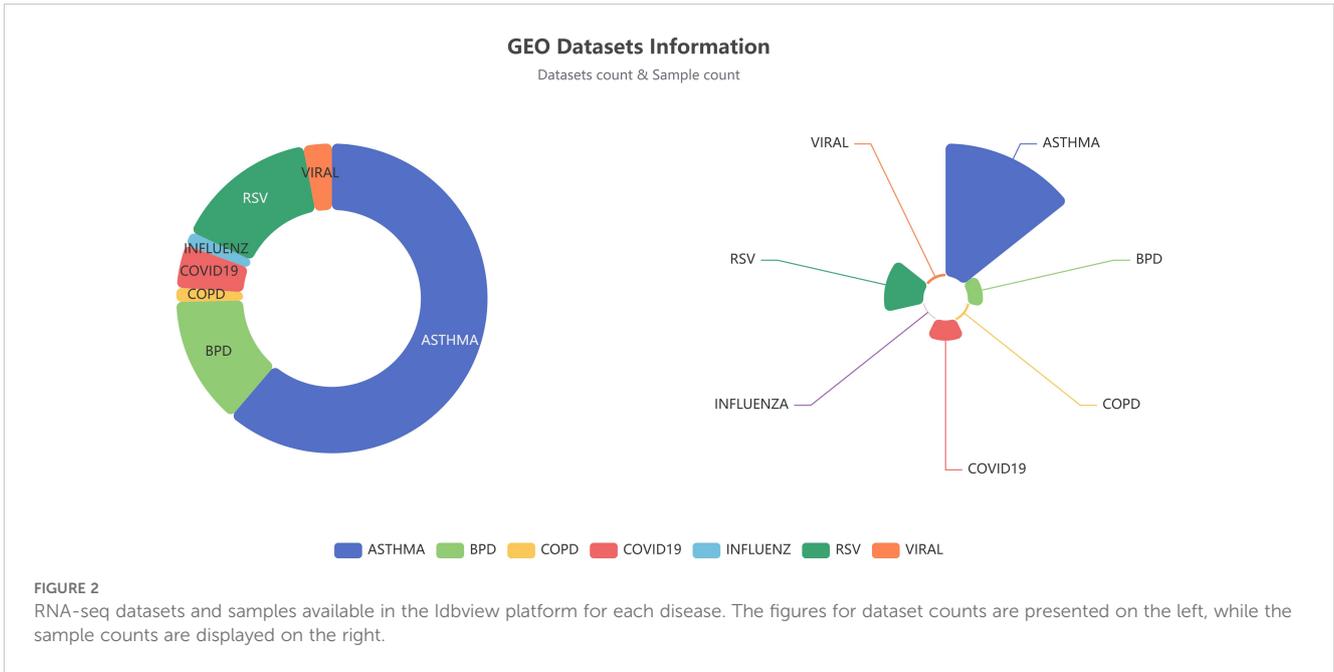
The omics data contains RNA-seq and scRNA modules, which focuses on RNA-seq sequencing data, and integrates tools such as differential gene expression analysis, DEG pathway enrichment analysis, and machine learning models such as logistic regression, RF, and SVM, etc. The RNA-seq module gathered 67 datasets

containing 11603 RNA-seq sequenced samples. Users can utilize these data for machine learning or differential gene analysis (Figures 2-4). Multiple download formats are available for users to choose from. The scRNA module introduced two respiratory single-cell shiny tools (Human-Lung v2 and SeuratV3Wizard) along with six respiratory single-cell datasets derived from the Lung Cell Atlas (18-23). These shiny tools enable users to analyze their own lung-related single-cell data. These datasets, integrated with corresponding webpage plug-in, facilitate data interaction at the cellular and genetic levels, enabling researchers to extract concise respiratory single-cell characteristics and comprehensive gene and cell information.

The GraphMed module provides analysis tools to aid users in analyzing their own data. For clinical data, users have access to various analysis and visualization tools, including correlation analysis, ANOVA, logistic regression, and global mapping. In the case of omics data, GraphMed integrates tools for differential gene expression analysis and DEG pathway enrichment. Additionally, basic visualization tools such as boxplots, violin plots, and bar charts utilizing ggplot2 are also available (Figure 5).

The Other module comprises other relevant features, such as pertinent information about “Our Lab” and “Hadv-Echart” (a specific human adenovirus data bar chart).

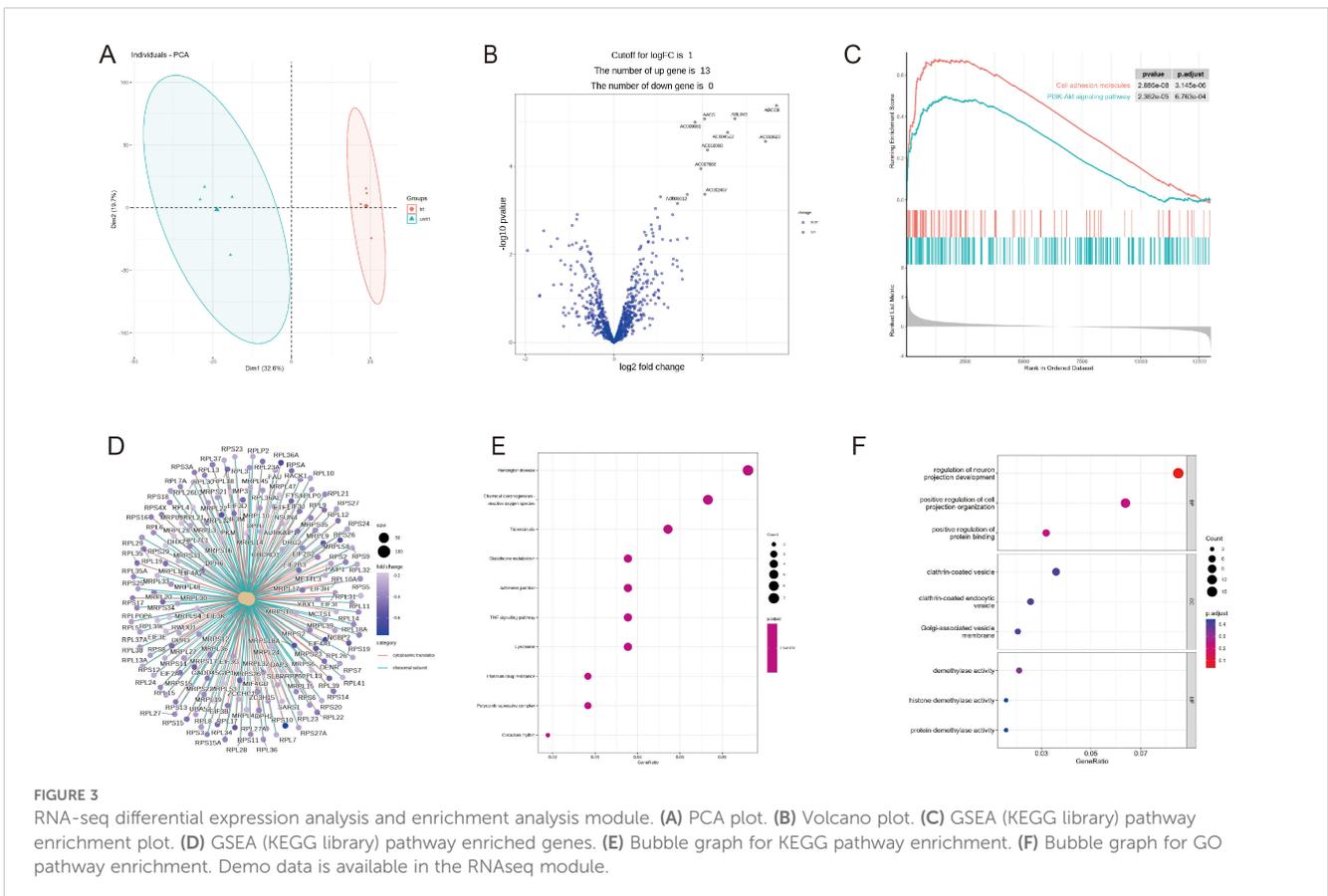
Through the consolidation of data and integration of analytical methods, we have enhanced the modules within the Idbview database, offering a robust resource to bolster research and applications targeted at respiratory diseases.

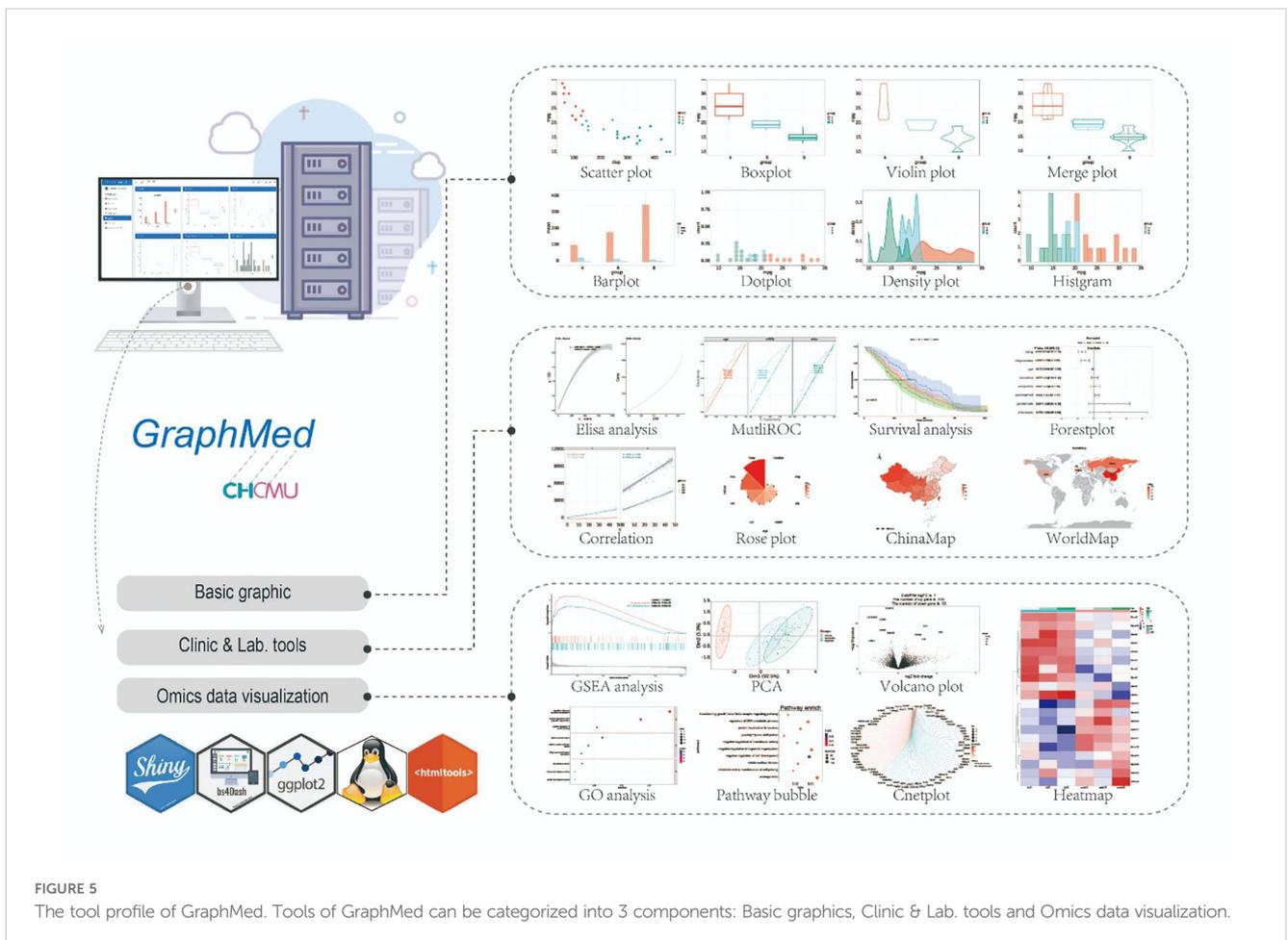
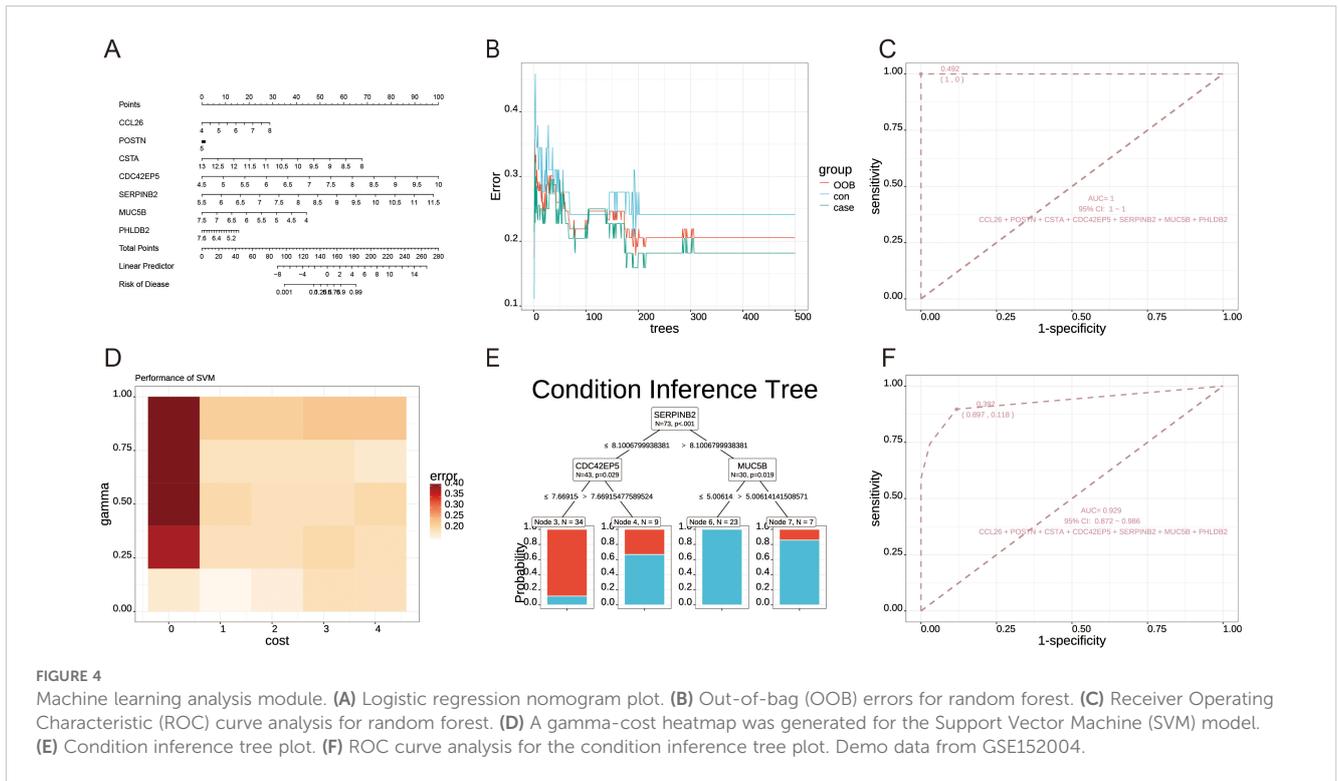


4 Discussion

As a novel database specialized in gathering information on respiratory diseases, Idbview will significantly aid researchers in their quest for disease knowledge. Idbview offers several advantages.

Firstly, the database comprises extensive and diverse respiratory disease data from various sources, catering to the practical needs of different researchers; in particular, it includes valuable clinical information from hospitals. Secondly, while some data are available in existing databases like GEO and WHO, Idbview





distinguishes itself through its user-friendly data processing and presentation, facilitating greater understanding and usability by individuals. Thirdly, Idbview incorporates logistic regression, conditional inference tree, RF, and SVM techniques, enabling researchers to analyze vast respiratory disease samples, identify potential biomarkers, and effectively explore the gene-disease relationships within large-scale sequencing data. Furthermore, Idbview offers over 30 common data analysis and visualization tools developed using R shiny, empowering users to analyze their data with ease, streamlining the analysis process, and ensuring high reproducibility. Significant limitations exist in this study. These limitations encompass incomplete data types, inability to merge omics data, limited inclusion of populations and diseases in GEO data (e.g., fibrotic lung diseases and rheumatoid arthritis-associated interstitial lung disease are not represented), a relatively slow pace of data incrementation owing to manual processing and uploading requirements, and restricted access to clinical samples due to ethical considerations and permissions. Moreover, the predominant inclusion of pediatric cases from CHCMU and GEO datasets may impede the generalizability of findings to adult respiratory diseases. Furthermore, the introduction of certain tools is currently only available in Chinese. Efforts will be made to expand the database by incorporating proteomics, metabolomics, and other data types, along with augmenting the clinical data collection.

Code availability

The website and database can be accessed through <https://idbview.com>. The R shiny tools code is available at <https://github.com/bingmp/idbview>, while the frontend code can be found at <https://github.com/bingmp/rVue>. The docker image source code has been uploaded to <https://hub.docker.com/r/pengbm/rshiny>.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Ethics statement

The studies involving humans were approved by Ethics Committee at Children's Hospital of Chongqing Medical University (2023 Ethics Committee (Research) No. 491). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

BP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. TL: Conceptualization, Data curation, Methodology, Resources, Software, Visualization, Writing – original draft. XF: Methodology, Software, Supervision, Writing – original draft. YZ: Investigation, Methodology, Software, Visualization, Writing – original draft. TW: Conceptualization, Funding acquisition, Validation, Writing – review & editing, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – original draft. ZF: Conceptualization, Funding acquisition, Software, Supervision, Writing – review & editing, Formal Analysis, Investigation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This project was supported by 82000034 from the National Natural Science Foundation of China, Chongqing Postdoctoral International Exchange Training Program(2021JLPY001) and National Clinical Medical Research Center (NCRC-2022-GP-08) during the conduct of the study.

Acknowledgments

We used Large language model ChatGPT-3.5 in the drafting of this paper for grammar and language refinement.

Conflict of interest

The authors declare 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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1460422/full#supplementary-material>

References

1. Mortality G B D, Causes of Death C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. (2016) 388:1459–544. doi: 10.1016/S0140-6736(16)31012-1
2. Societies F. *The global impact of respiratory disease*. European Respiratory Society (2017).
3. Members C-N, Partners. Database resources of the national genomics data center, China national center for bioinformation in 2024. *Nucleic Acids Res.* (2024) 52:D18–32. doi: 10.1093/nar/gkad1078
4. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* (2013) 41:D991–5. doi: 10.1093/nar/gks1193
5. Cantelli G, Cochrane G, Brooksbank C, McDonagh E, Flicek P, McEntyre J, et al. The European Bioinformatics Institute: empowering cooperation in response to a global health crisis. *Nucleic Acids Res.* (2021) 49:D29–37. doi: 10.1093/nar/gkaa1077
6. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
7. Li J, Miao B, Wang S, Dong W, Xu H, Si C, et al. Hiplot: a comprehensive and easy-to-use web service for boosting publication-ready biomedical data visualization. *Brief Bioinform.* (2022) 23. doi: 10.1093/bib/bbac261
8. Wu T, Hu E, Ge X, Yu G. nCov2019: an R package for studying the COVID-19 coronavirus pandemic. *PeerJ.* (2021) 9:e11421. doi: 10.7717/peerj.11421
9. Ren Y, Yu G, Shi C, Liu L, Guo Q, Han C, et al. Majorbio Cloud: A one-stop, comprehensive bioinformatic platform for multiomics analyses. *iMeta.* (2022) 1:e12. doi: 10.1002/imt.v1.2
10. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas. *Elife.* (2017) 6:e27041. doi: 10.7554/eLife.27041
11. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E, et al. An integrated cell atlas of the lung in health and disease. *Nat Med.* (2023) 29:1563–77. doi: 10.1038/s41591-023-02327-2
12. Granjon D. *bs4Dash: A 'Bootstrap 4' Version of 'shinydashboard'*. (2024).
13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
15. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag (2016).
16. Allen J. *htmltools: Tools for HTML*. (2023).
17. Sievert C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC (2020).
18. He P, Lim K, Sun D, Pett JP, Jeng Q, Polanski K, et al. A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell.* (2022) 185:4841–60.e25. doi: 10.1016/j.cell.2022.11.005
19. Madissoon E, Oliver AJ, Kleshchevnikov V, Wilbrey-Clark A, Polanski K, Richoz N, et al. A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat Genet.* (2023) 55:66–77. doi: 10.1038/s41588-022-01243-4
20. Lim K, Donovan APA, Tang W, Sun D, He P, Pett JP, et al. Organoid modeling of human fetal lung alveolar development reveals mechanisms of cell fate patterning and neonatal respiratory disease. *Cell Stem Cell.* (2023) 30:20–37.e9. doi: 10.1016/j.stem.2022.11.013
21. Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, Simon LM, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med.* (2019) 25:1153–63. doi: 10.1038/s41591-019-0468-5
22. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.* (2019) 21:1. doi: 10.1186/s13059-019-1906-x
23. Barnes JL, Yoshida M, He P, Worlock KB, Lindeboom RGH, Sui C, et al. Early human lung immune cell development and its role in epithelial cell fate. *Sci Immunol.* (2023) 8:eadf9988. doi: 10.1126/sciimmunol.adf9988