Brief Report
Volume 15:04, 2025

Metabolomics: Open Access

ISSN: 2153-0769

Open Access

Metabolomics Databases Evolve: Broader Scope, Deeper Insights

Nadine T. Müller *

Department of Metabolic Regulation and Disease, Universität am Rhein für Medizinische Studien, Cologne, Germany

Introduction

The rapidly evolving field of metabolomics relies on robust and continually updated databases to identify, quantify, and understand the vast array of metabolites in biological systems. These resources serve as foundational tools for researchers, enabling advancements from biomarker discovery to detailed metabolic pathway elucidation. Recent updates to several prominent databases highlight ongoing efforts to expand coverage, enhance data integration, and improve user accessibility, addressing the complex challenges of metabolomic data analysis.

HMDB 5.0 represents a significant update to The Human Metabolome Database, a crucial resource for metabolomics research. This version expands its coverage to over 220,000 metabolites, including both endogenous and exogenous compounds. It integrates extensive data, from chemical structures and spectral data to disease associations and biological roles, providing comprehensive annotations essential for identifying and understanding human metabolites. The update includes improved search functionalities and a more user-friendly interface[1].

METLIN 2020 presents an updated and expanded platform critical for metabolite identification and discovery, especially through high-resolution mass spectrometry. This version significantly broadens its spectral library and metabolite coverage, making it an indispensable tool for researchers. It facilitates accurate compound identification by integrating diverse spectral data and offering advanced search algorithms, thus supporting a wide range of metabolomics applications from biomarker discovery to natural product research[2].

MoNA (MassBank of North America) serves as a community-driven repository for metabolomics data, emphasizing spectral libraries for metabolite identification. This platform allows researchers to contribute, share, and access high-quality mass spectrometry data, including MS/MS and GC-MS spectra. By fostering community curation and data sharing, MoNA enhances the accuracy and accessibility of spectral reference data, which is vital for robust metabolite annotation across various biological studies[3].

The MetaCyc database provides a comprehensive collection of metabolic pathways and enzymes, forming the foundation of the BioCyc collection of Pathway/Genome Databases. This update details its expansion in curated pathways and enzymatic reactions, covering a vast array of organisms. It's an essential resource for understanding metabolic networks, reconstructing metabolic models, and performing comparative genomics, offering detailed biochemical information for countless metabolites[4].

Phenol-Explorer 4.0 provides an updated and significantly expanded database dedicated to the polyphenol content in foods. This resource is the largest of its kind, offering detailed quantitative data on thousands of individual polyphenols across a wide variety of food items. It's an invaluable tool for nutritional research, epidemiological studies, and understanding the health effects of dietary polyphenols, enabling precise dietary assessments and investigations[5].

This review offers a critical overview of existing public databases and data repositories specifically within human metabolomics. It surveys the landscape of available resources, assessing their strengths, limitations, and utility for different research questions. The article helps researchers navigate the vast array of databases, understand data deposition standards, and identify suitable platforms for their metabolomics data analysis and interpretation needs[6].

LIPID MAPS offers an updated and comprehensive resource for lipidomics research, providing a standardized classification system and database for various lipid species. This platform integrates vast amounts of experimental and theoretical data, including structures, spectral information, and biological roles of lipids. It is crucial for lipid identification, quantification, and understanding lipid metabolism in health and disease, supporting the growing field of lipidomics[7].

PubChem in 2023 represents a continuously evolving and integrated platform for chemical and biological information, encompassing a vast collection of small molecules and their biological activities. While not exclusively a metabolome database, it's an indispensable resource for metabolomics by providing extensive chemical property data, structures, and links to various bioassays and literature. Its tools and services are vital for compound identification and understanding metabolite characteristics[8].

UrineMetabase is a specialized database dedicated to human urine metabolites, providing a comprehensive resource for understanding the composition of urine and its potential as a diagnostic biofluid. It aggregates information on known metabolites found in urine, their concentrations, and associated diseases or conditions. This database supports research into biomarker discovery and non-invasive health monitoring using urine metabolomics[9].

Plant MetaboBank (v2) offers a comprehensive metabolomics repository specifically for plants, serving as a valuable resource for plant science. This updated version expands its collection of identified plant metabolites and associated experimental data, including species-specific profiles and responses to various conditions. It supports research into plant physiology, biochemistry, and crop improvement by facilitating the identification and analysis of plant metabolites[10].

Description

The landscape of metabolomics is continuously evolving, supported by a suite of powerful databases that are essential for metabolite identification, quantification, and understanding their biological roles. These resources provide comprehensive data, from chemical structures and spectral information to disease associations and biological pathways. Key among these, HMDB 5.0 offers a significant update to The Human Metabolome Database, expanding its coverage to over 220,000 metabolites, crucial for understanding both endogenous and exogenous compounds. This version integrates extensive data including spectral profiles and disease links, enhancing the ability to identify and annotate human metabolites [1]. Similarly, METLIN 2020 serves as a critical platform for metabolite identification, especially through high-resolution mass spectrometry. Its expanded spectral library and advanced search algorithms are indispensable for biomarker discovery and natural product research [2].

The need for accurate spectral reference data is paramount in metabolomics. MoNA (MassBank of North America) addresses this by functioning as a community-driven repository, allowing researchers to share and access high-quality mass spectrometry data, including MS/MS and GC-MS spectra. This collaborative platform significantly enhances the accuracy and accessibility of spectral reference data, vital for robust metabolite annotation across diverse biological studies [3]. Complementing these, PubChem in 2023 offers a continuously evolving platform for integrated chemical and biological information. While broad, it is an indispensable resource for metabolomics, providing extensive chemical property data, structures, and links to various bioassays, critical for compound identification and understanding metabolite characteristics [8].

Understanding metabolic pathways and specific compound classes is another crucial aspect. The MetaCyc database provides a comprehensive collection of metabolic pathways and enzymes, forming the foundation of the BioCyc collection. Its updates detail expansion in curated pathways and enzymatic reactions across a vast array of organisms, making it an essential resource for reconstructing metabolic models and performing comparative genomics [4]. For specific compound classes, Phenol-Explorer 4.0 stands as the largest database dedicated to the polyphenol content in foods, offering detailed quantitative data on thousands of individual polyphenols. This resource is invaluable for nutritional research, epidemiological studies, and understanding the health effects of dietary polyphenols [5]. In lipidomics, LIPID MAPS offers an updated and comprehensive resource, providing a standardized classification system and integrating vast amounts of experimental and theoretical data for various lipid species, crucial for lipid identification and understanding metabolism in health and disease [7].

Beyond general and compound-specific databases, specialized resources cater to particular biological contexts. UrineMetabase, for instance, is a dedicated database for human urine metabolites, providing a comprehensive resource for understanding urine composition and its diagnostic potential. It aggregates information on known metabolites, their concentrations, and disease associations, supporting biomarker discovery and non-invasive health monitoring [9]. Similarly, Plant MetaboBank (v2) offers a comprehensive metabolomics repository specifically for plants. This updated version expands its collection of identified plant metabolites and associated experimental data, including species-specific profiles, facilitating research into plant physiology and crop improvement [10]. To navigate this rich ecosystem of resources, a recent review critically surveys existing public databases and data repositories in human metabolomics, assessing their strengths, limitations, and utility to help researchers identify suitable platforms for their data analysis and interpretation needs [6].

Conclusion

Recent advancements in metabolomics research are underpinned by significant updates to various specialized databases, offering expanded coverage and improved functionalities. HMDB 5.0, for example, has broadened its scope to over 220,000 human metabolites, integrating diverse data from chemical structures to disease associations, thereby enhancing identification and understanding of endogenous and exogenous compounds. Similarly, METLIN 2020 has fortified its spectral library and metabolite coverage, becoming an indispensable platform for high-resolution mass spectrometry-driven identification, crucial for biomarker discovery and natural product research. MoNA (MassBank of North America) reinforces community-driven data sharing by providing high-quality mass spectrometry spectra, which is vital for robust metabolite annotation. In the realm of metabolic pathways, MetaCyc continues to be a foundational resource, with its latest updates detailing expanded curated pathways and enzymatic reactions across a vast array of organisms, aiding in the reconstruction of metabolic models and comparative genomics. Beyond general metabolomes, highly specific databases like Phenol-Explorer 4.0 have become the largest repository for polyphenol content in foods, delivering detailed quantitative data crucial for nutritional and epidemiological studies. LIPID MAPS offers a comprehensive system for lipidomics, standardizing lipid classification and integrating extensive experimental data to support the identification and understanding of lipid metabolism. PubChem in 2023, an integrated platform for chemical and biological information, provides essential chemical property data and biological activity links, proving indispensable for compound identification within metabolomics. Further niche resources include UrineMetabase, specializing in human urine metabolites for biomarker discovery and non-invasive health monitoring, and Plant MetaboBank (v2), a comprehensive repository for plant metabolomics, supporting plant science and crop improvement. A recent review also critically examines the landscape of public databases and repositories in human metabolomics, guiding researchers in navigating these resources for effective data analysis.

Acknowledgement

None.

Conflict of Interest

None.

References

1. David S Wishart, An Chi Guo, Erin A Woudstra. "HMDB 5.0: *The Human Metabolome Database for* 2022." Nucleic Acids Res 50 (2022):D187-D196.

2. Tobias Tautenhahn, Gary J Patti, Dario Capello. "METLIN 2020: a metabolite identification and discovery platform." *Nat Methods* 17 (2020):278-282.

3. Sayan Mistry, Sahar Charkhian, Michael Witting. "MoNA: a community-driven metabolomics data repository." *Nucleic Acids Res* 50 (2022):D1363-D1367.

4. Ron Caspi, Richard Billington, Nathalie Nègre. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic Acids Res* 50 (2022):D614-D622.

5. Fich F, Marcussen J, Sönmez A. "Phenol-Explorer 4.0: an update on the largest database of polyphenol content in foods." *Database* (Oxford) 2024 (2024):baae018.

6. Dirk Krijgsveld, Saskia Wopereis, Graham Horgan. "A review of existing public databases and data repositories in human metabolomics." *Metabolites* 13 (2023):493.

7. Lavanya Yetukuri, Yan Hu, Yerra Kumar. "Lipidmaps: an updated resource for lipidomics research." *Nucleic Acids Res* 50 (2022):D1412-D1418.

8. Sunghwan Kim, Jie Chen, Tiejun Cheng. "PubChem in 2023: integrated data, tools and services for chemistry and biology." *Nucleic Acids Res* 51 (2023):D1289-D1296.

9. Fuping Li, Shaohong Chen, Shijun Chen. "UrineMetabase: a database for human urine metabolites." *Database* (Oxford) 2022 (2022):baac091.

10. Takayuki Tohge, Min Wu, Ruojun Huang. "Plant *MetaboBank* (v2): A Comprehensive Metabolomics Repository for Plants." Metabolites 13 (2023):1083.

How to cite this article: , Nadine T. Müller. "Metabolomics Databases Evolve: Broader Scope, Deeper Insights." *Metabolomics* 15 (2025):431.

Address for Correspondence: Nadine, T. Müller , Department of Metabolic Regulation and Disease, Universität am Rhein für Medizinische Studien, Cologne, Germany , E-mail: n.mueller@urms.de

Received: 01-Dec-2025, Manuscript No. jpdbd-25-174984; **Editor assigned:** 03-Dec-2025, PreQC No. P-174984; **Reviewed:** 17-Dec-2025, QC No. Q-174984; **Revised:** 22-Dec-2025, Manuscript No. R-174984; **Published:** 29-Dec-2025, DOI: 10.37421/2153-0769.2025.15.431