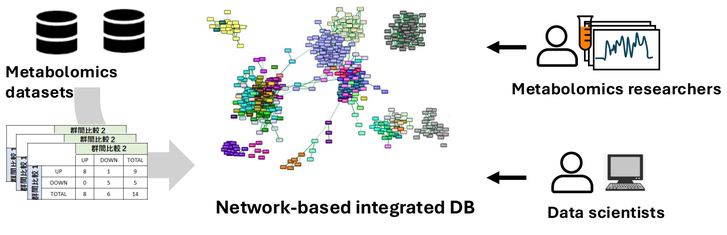
大規模メタボローム統合解析プラットフォームintegMET

早川英介[1,2]、西田孝三[1]、高橋みき子[1]、岡昂輝[3]、松田りら[4,5]、山本博之[5]、津川裕司[3]、河野信[6]

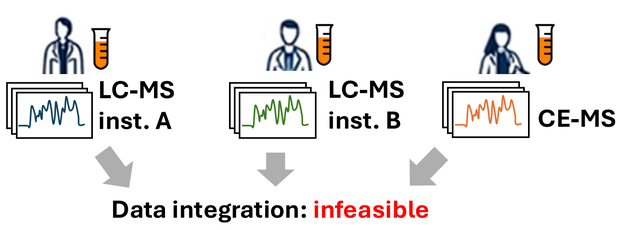
1 九州工業大学、2 理化学研究所、3 東京農工大学、4 慶應義塾大学、5 ヒューマン・メタボローム・テクノロジーズ株式会社、6 北里大学

Summary

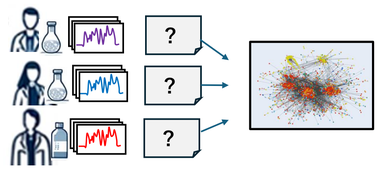


Metabolomics has advanced significantly, yielding extensive datasets stored in data repositories. However, the full potential of these data is limited by difficulties in integrating and reanalyzing information from different studies due to variations in analytical settings. We have been developing a network-based data integration platform integMET, to reanalyze cross-study metabolomics data. By quantifying similarities in differential metabolite profiles and analyzing textual metadata, we create a unified network linking metabolic data and scientific concepts. This method enhances data access and promotes cross-disciplinary scientific advancement.

Background



In mass spectrometry-based metabolomics, variability in instrumentation and analytical condition poses significant challenges for standardization of metabolomics data. Differences in absolute quantification and instrument-dependent signal intensities hinder data comparability. Additionally, metabolite detection consistency is impacted by variations in instrumental settings, affecting cross-platform data reliability.



Metabolomics data repositories, such as MetaboLights and Metabolomics Workbench, contain a vast amount of data derived from diverse samples, experimental conditions, and research backgrounds. However, the metadata within these repositories is often inadequate and lacks standardization. This lack of comprehensive and uniform metadata makes the reanalysis and integration of data across networks challenging.

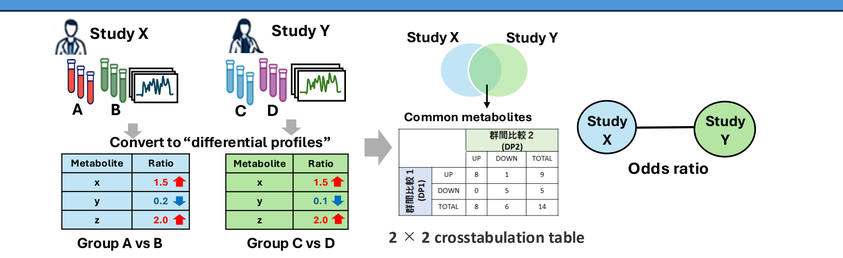
Methods

By addressing the challenges in metabolomics data integration, we are currently developing a data integration platform for cross-study reanalysis.

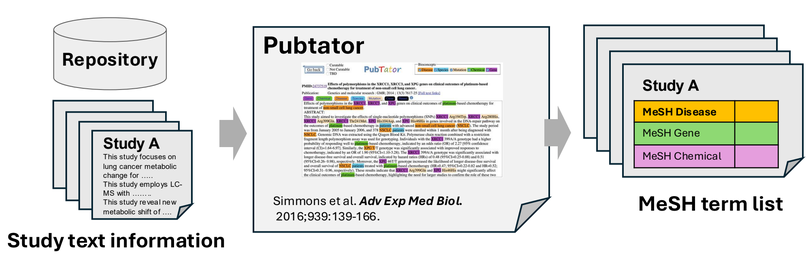
1.Metabolite differential profile-based Integration: Employing iDMET (Matsuta et al ***BMC Bioinformatics*** 2022) a differential profile-based approach to integrate metabolomics data from various studies.

2.Metadata-Based Integration: Keyword-based data integration to support interpretation of complex metabolome data sets.

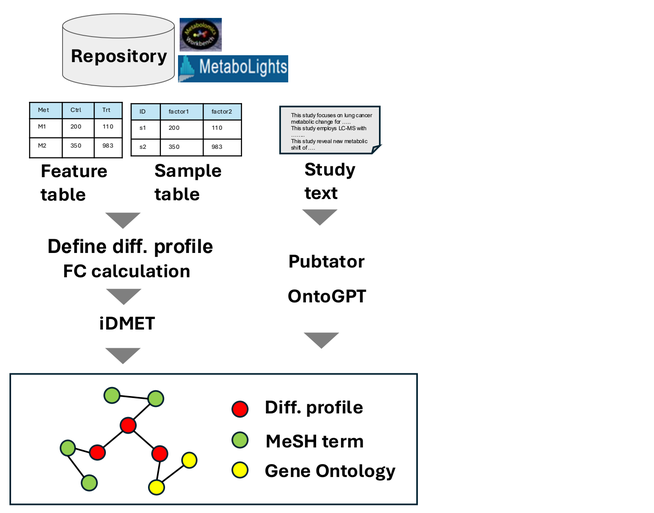
3.Integrated Network for Data Exploration: Developing an integrated network platform that facilitates the exploration of vast metabolomics data, paving the way for novel discoveries and insights.



1. Metabolite differential profile-based integration

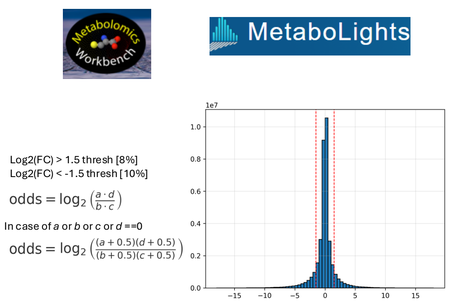


2. Metadata-Based Integration

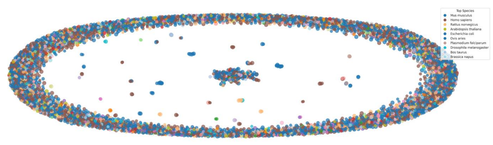


3. Integrated Network for Data Exploration

Results and Discussion



Distribution of $\log_2$ Fold-Change Values Across All Differential Profiles



integMET raw graph

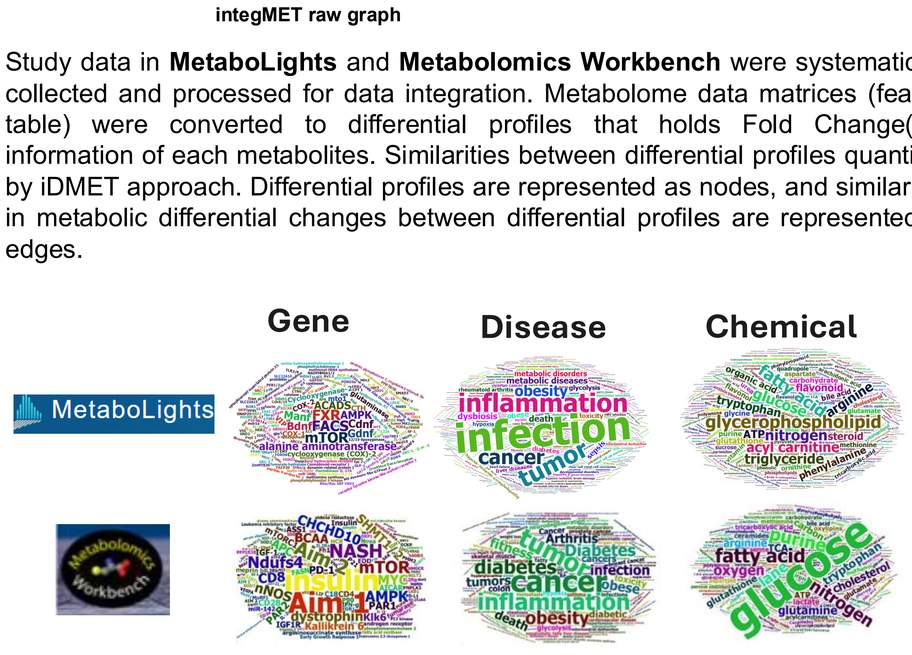
2,287 studies
- MetaboLights[ML]: 511 (2023 Jun)
- Metabolomics Workbench [MW]: 1,776 (2025)

68,733 unique InChIKeys
- ML: 54,744
- MW: 29,114
- [intersection]: 15,125

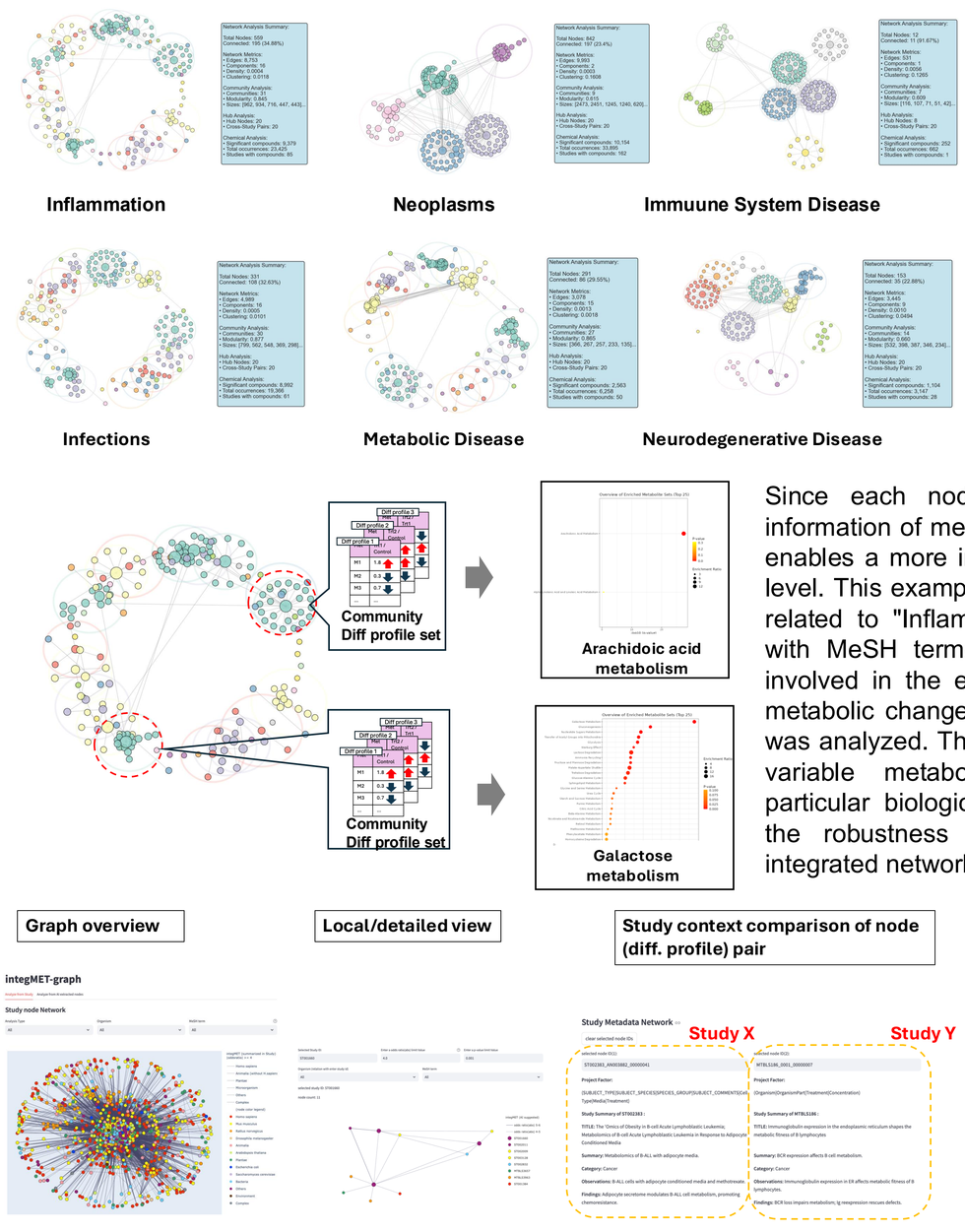
440,299 diff profiles
- ML: 111,584
- MW: 328,715

36,028,171 FCs
- ML: 11,722,604
- MW: 24,305,567

Study data in **MetaboLights** and **Metabolomics Workbench** were systematically collected and processed for data integration. Metabolome data matrices (feature table) were converted to differential profiles that holds Fold Change(FC) information of each metabolites. Similarities between differential profiles quantified by iDMET approach. Differential profiles are represented as nodes, and similarities in metabolic differential changes between differential profiles are represented as edges.



Study descriptions acquired from repositories were processed with PubTator to extract key biological entities (MeSH terms) such as diseases, genes, chemicals, and species. The extracted information serves as a valuable resource for assessing the data content and diversity within the repository.

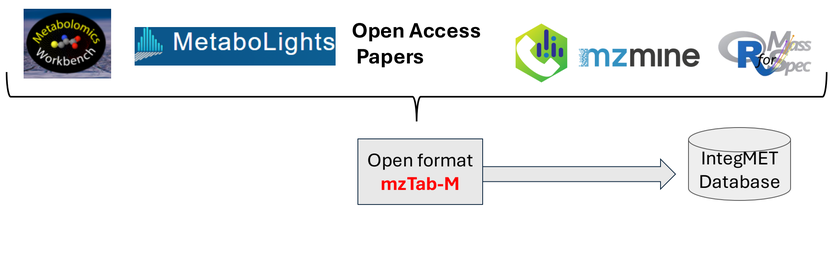


By leveraging graph data structured at both the metabolite and metadata levels, repository-wide integrated metabolomics data can be reanalyzed flexibly and efficiently. Ontology (MeSH term and GO)-based filtering enables visualization of the distribution of studies related to various subjects and their similarity in metabolic dynamics.

Since each node and edge carries fold-change information of metabolites from individual studies, this enables a more in-depth reanalysis at the metabolite level. This example shows extraction of a subnetwork related to "Inflammation" using metadata annotated with MeSH terms. The composition of metabolites involved in the edges, representing the similarity in metabolic changes, within the extracted subnetworks was analyzed. This analysis provides insights into the variable metabolites highly associated with a particular biological entity (inflammation), confirming the robustness and biological relevance of the integrated network.

This data integration platform is currently under development, along with a web browser-based user interface. The platform supports large-scale data reanalysis, facilitating the exploration of data to identify novel findings, such as common metabolic changes across different biological backgrounds

Perspectives and Future Work



We are currently developing the interface for the integrated dataset, with plans to (beta) release it by the end of the current fiscal year. Concurrently, to enable comprehensive integration of metabolomics research sources beyond traditional repositories, we have adopted the community-standard open format mzTab-M for system-wide data management. We will continue to advance this database’s development to contribute to the fields of metabolomics and bioinformatics.

This research is supported by the JST-NBDC Database Integration Coordination Program under Grant Number JPMJND2403