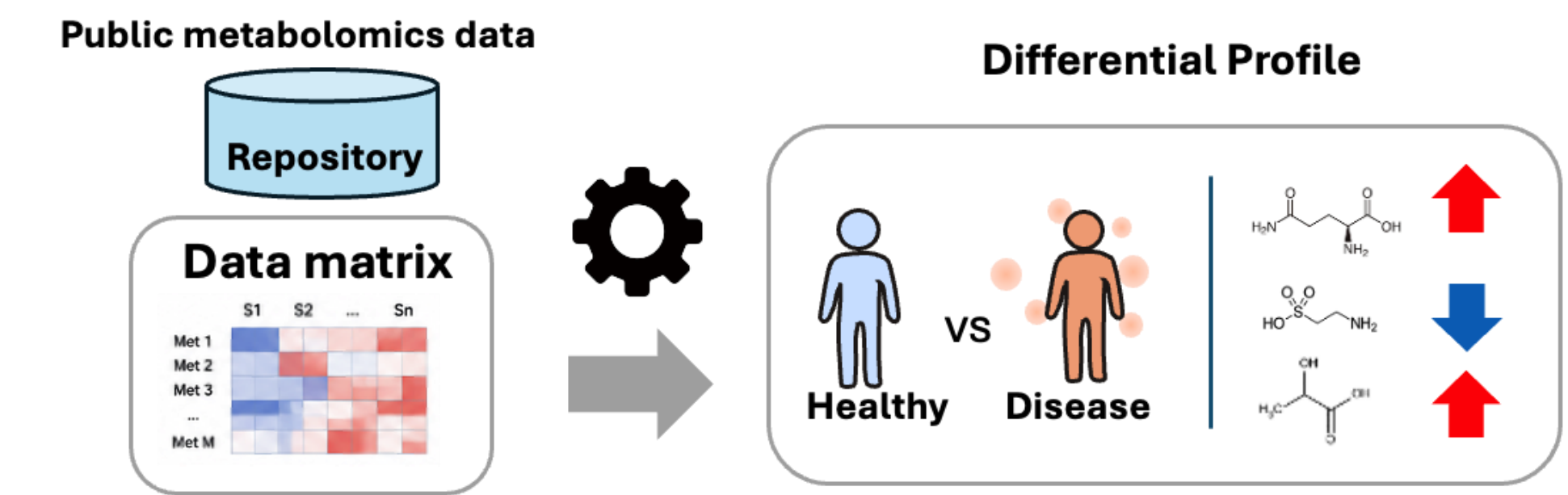


integMET:公共メタボロミクスデータの統合的再解析基盤の構築

早川英介^{1,2}、西田孝三²、高橋みき子²、古賀優将¹、松田りら^{4,5}、山本博之⁵、河野信⁶

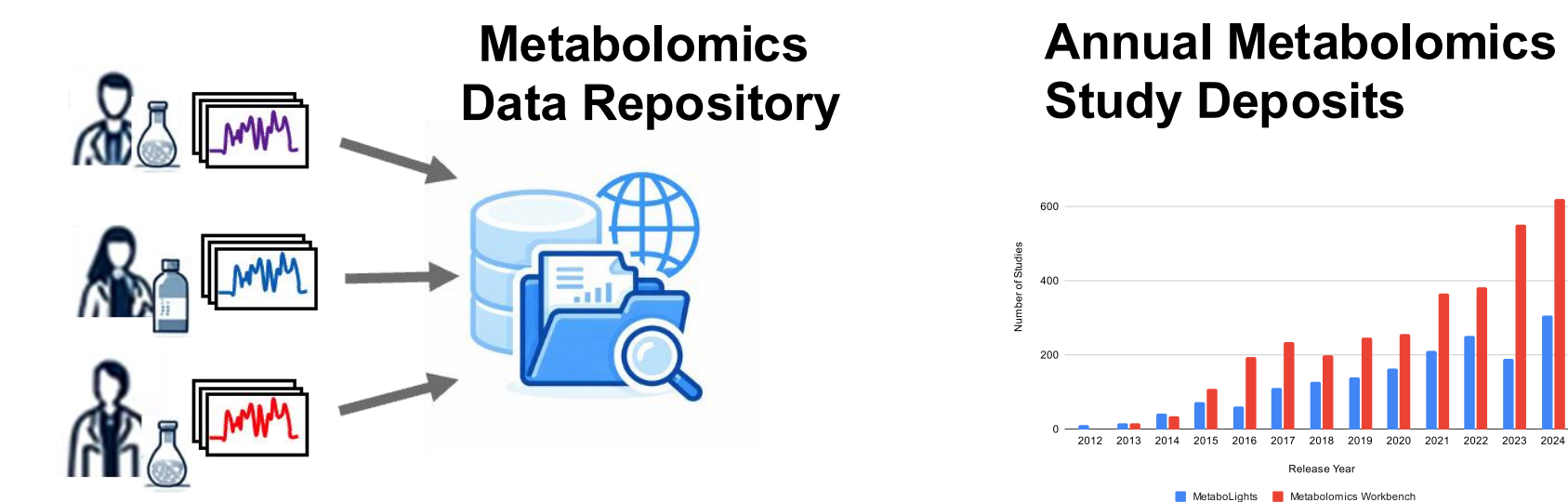
1 九州工業大学、2 理化学研究所、4 慶應義塾大学、5 ヒューマン・メタボローム・テクノロジーズ、6 北里大学

Summary

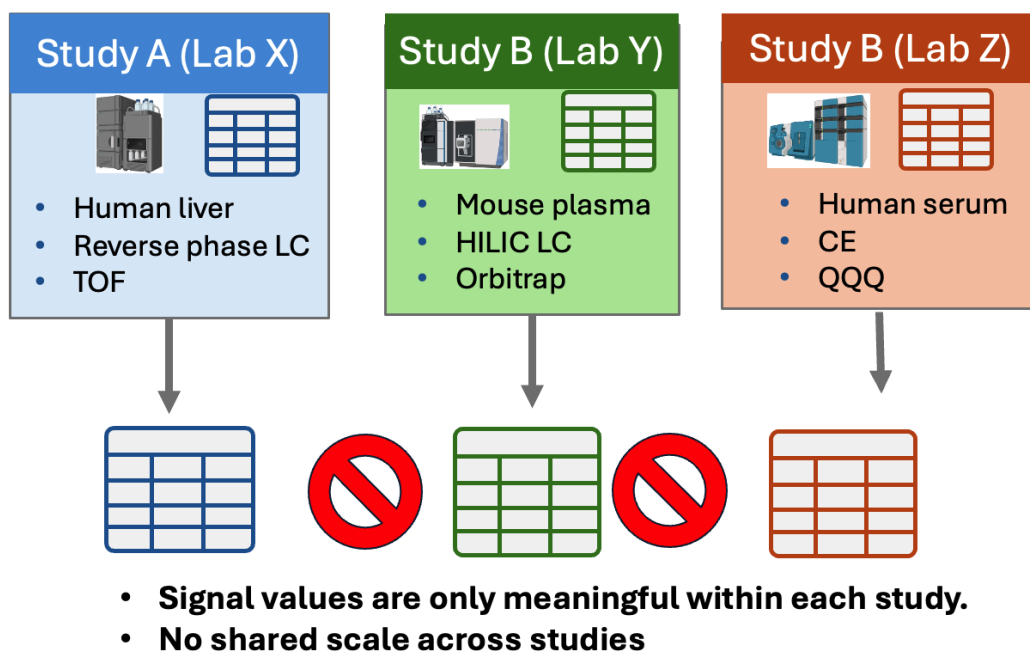


Metabolomics has produced extensive datasets now archived in public repositories, yet much of their value remain largely unused. Differences in analytical conditions prevent direct comparison of signal intensities mass spectrometry signals across studies, and incomplete metadata makes relevant datasets hard to find. To unlock these data, we are building a database that recasts each study as group-to-group comparisons, a Differential Profiles expressed as fold changes of metabolites observed in group-to-group comparisons, rather than as absolute intensities of raw data matrices. This representation makes measurements comparable across studies and enables cross-study exploration from either a metabolite or a study-background angle, promoting data reuse and cross-disciplinary discovery.

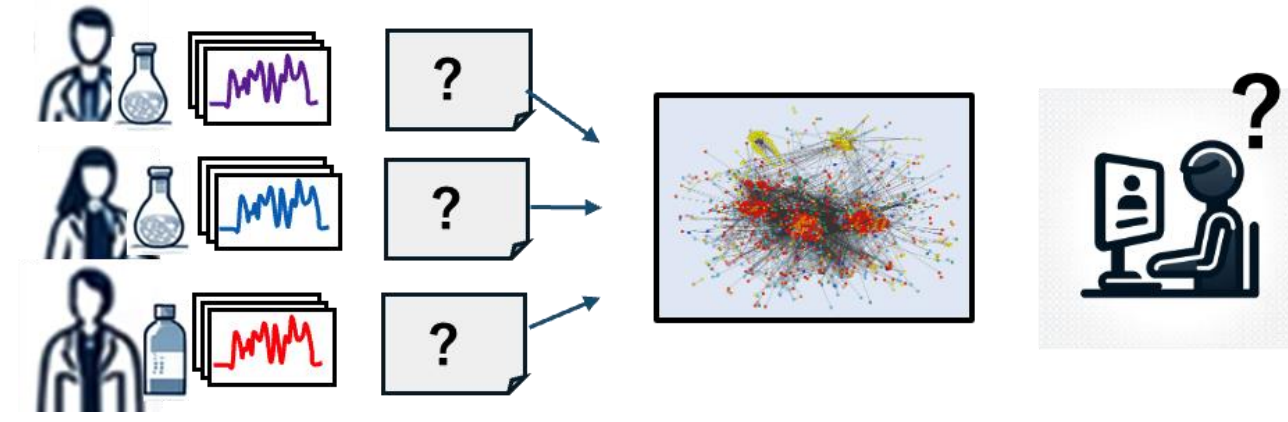
Background



Mass spectrometry-based metabolomics is widely used, and its data continue to accumulate in public repositories such as MetaboLights and Metabolomics Workbench.



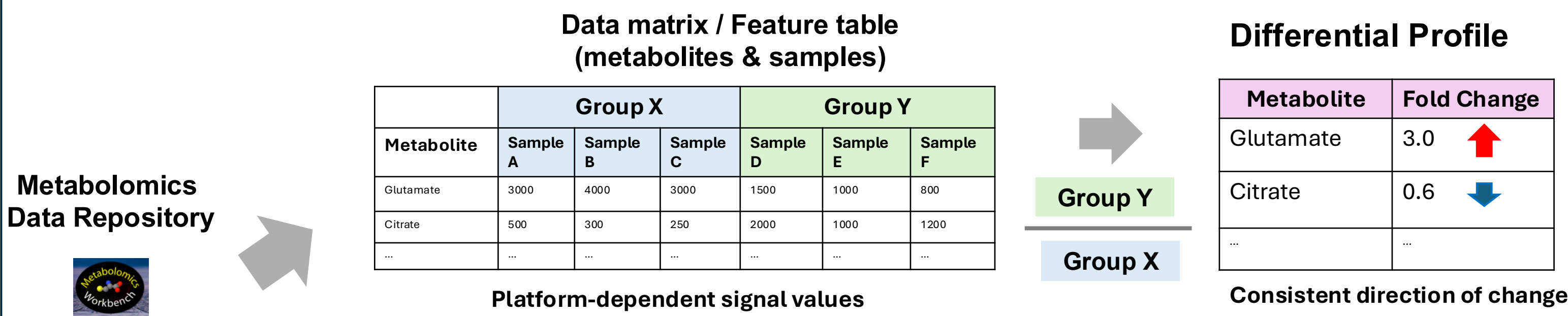
However, the intensity values of mass spectrometry measurements are instrument- and condition-dependent and therefore meaningful only within a single study. Therefore signal intensities are not compared to other projects, preventing pan-repository, cross study analysis.



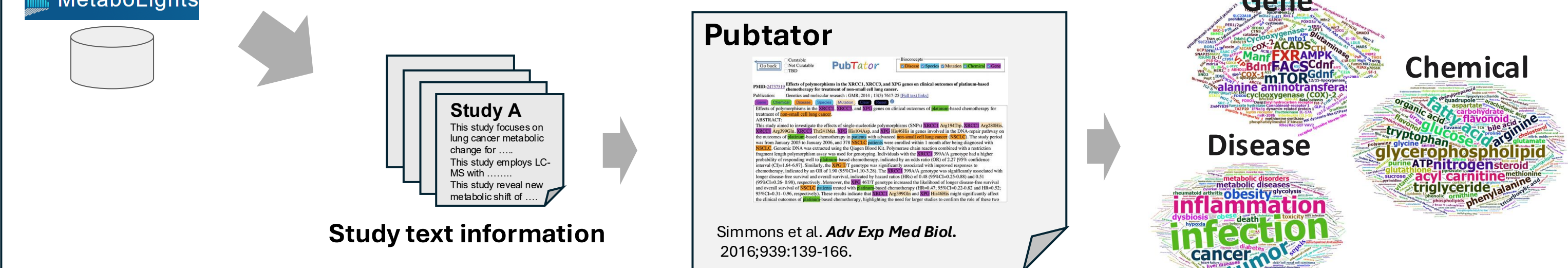
Moreover, the metadata accompanying repository entries are often incomplete and non-standardized, making even the discovery of relevant studies difficult. As a result, the vast body of metabolomics data already deposited cannot be reused or integrated, leaving most of its potential value unrealized.

Methods

1. Generate “Differential Profile”



2. Extract Biological meaning/category



Study

- Raw data matrices
- Sample information
- Study-level metadata

Differential Profile

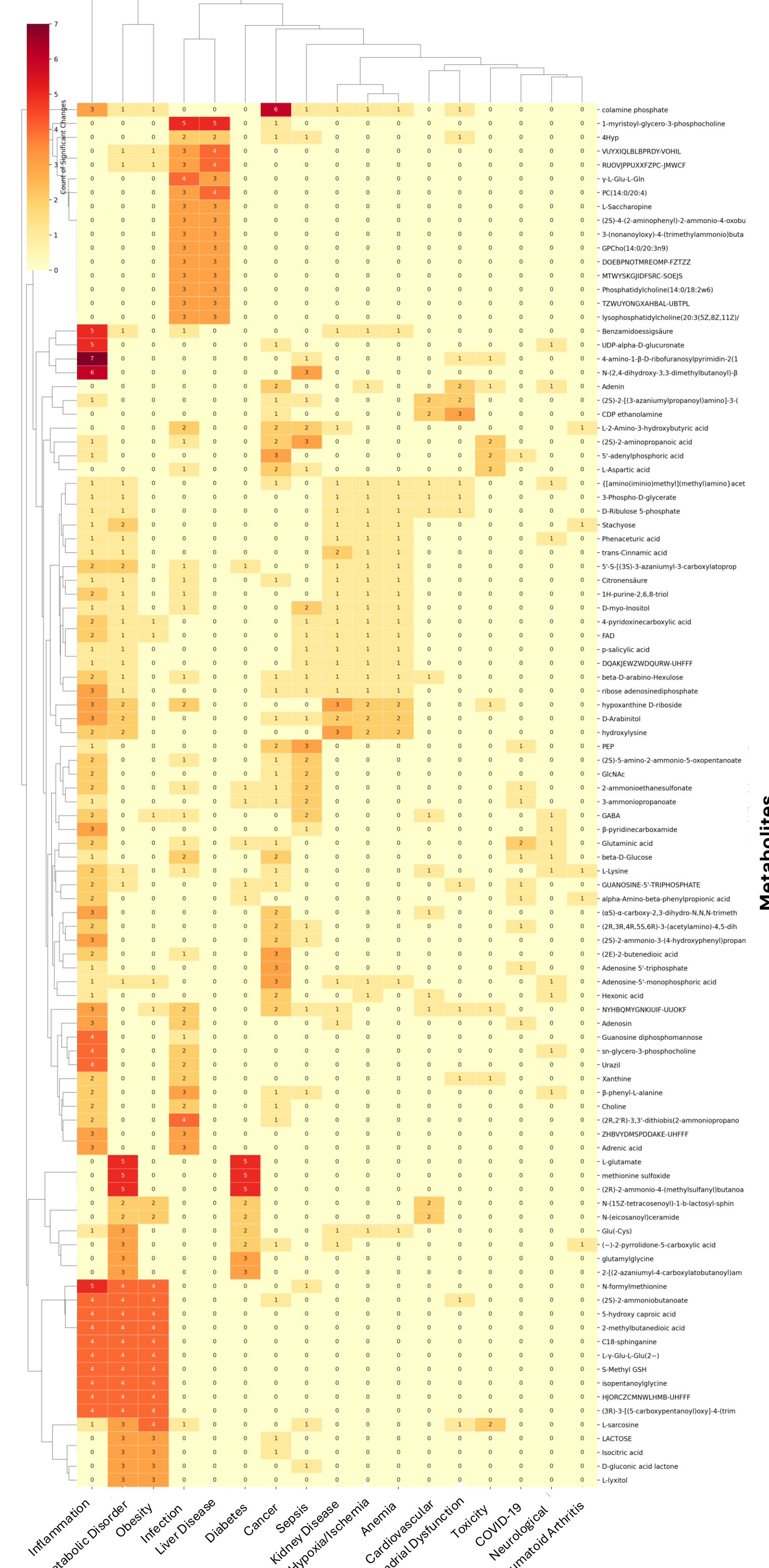
- Biological comparisons
- Metabolic changes

Metabolite

- Molecular identity
- Associated biological comparison

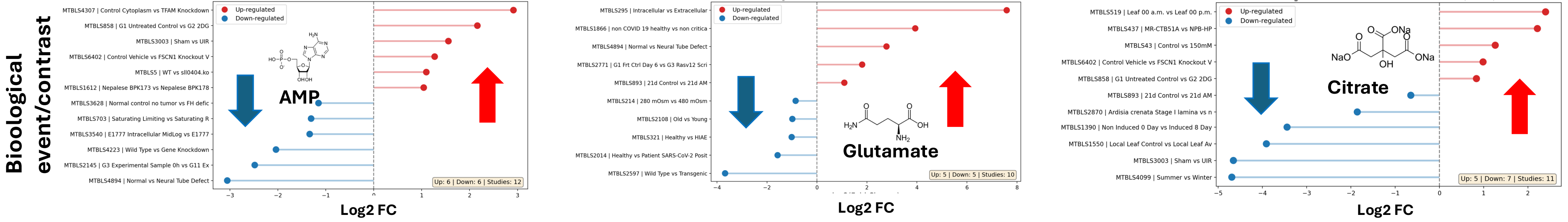
At its core, the database reorganizes repository data around the “Differential Profile” structure into three linked layers. The Study layer captures each dataset as a source, with its research background and sample context. The Differential Profile layer captures the biological contrasts within a study together with the metabolite increases and decreases that accompany them. The Metabolite layer holds each compound with annotations such as the contrasts in which it varies and its pathway membership. Linking these three layers turns heterogeneous repository data into a single multi-layer resource that can be entered from any point and traversed to the others.

Results

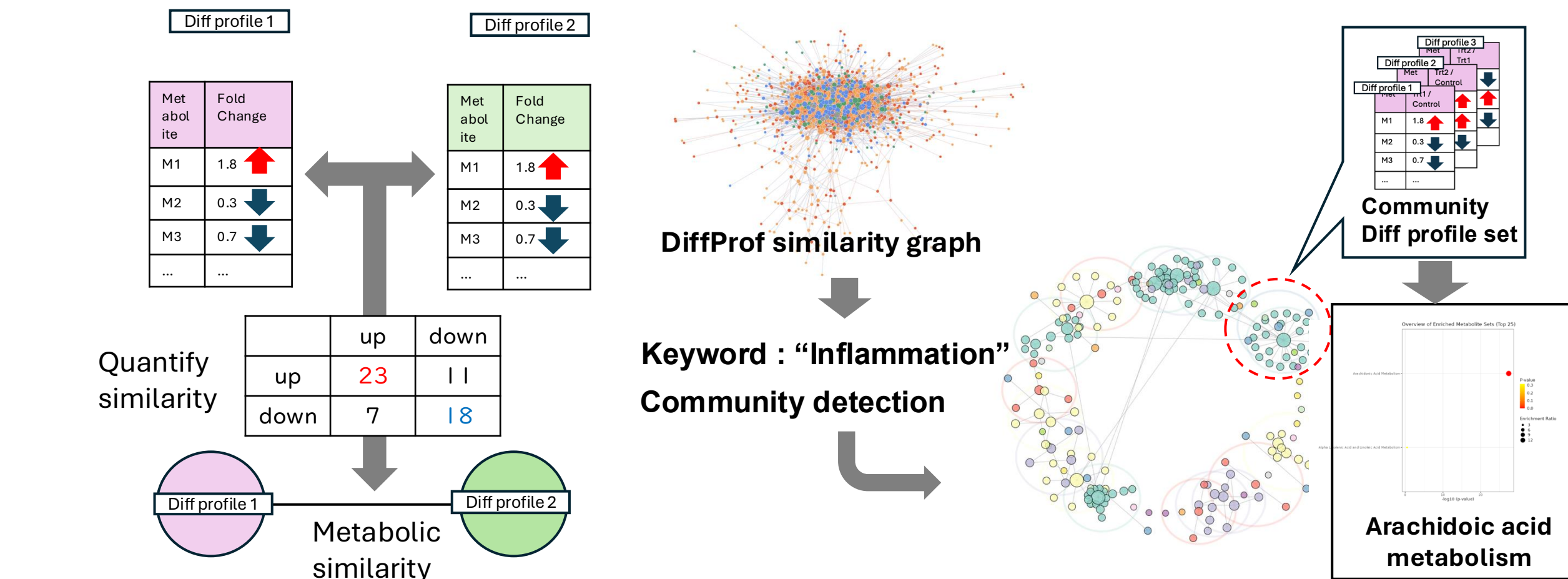


Metabolite-disease association map

Because every dataset is expressed on a common scale, the differential profiles become directly comparable across studies. As one application, all disease-related differential profiles were extracted and summarized as a disease-by-metabolite heatmap, in which each cell is colored by how often that metabolite is found to vary: redder cells mark disease-metabolite pairs whose variation is observed in more studies. These high-frequency cells stand out, revealing cross-study trends in disease-associated metabolic variation that are invisible within any single dataset.

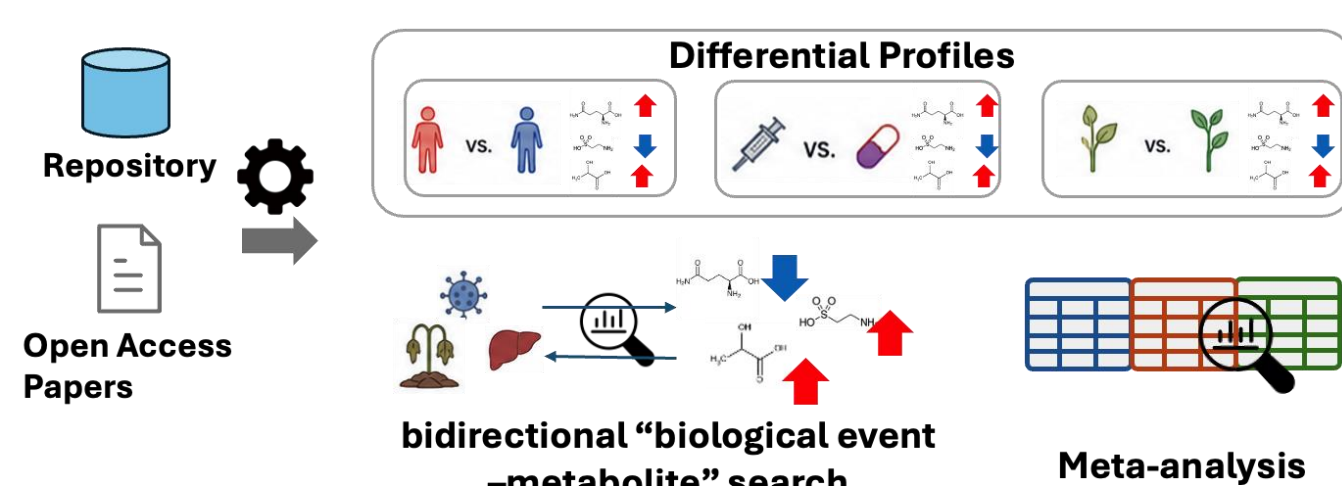


Because all Differential Profiles share common metabolite identifiers, any compound or condition can be queried across the entire collection derived from all studies in repositories. For a given metabolite, its behavior across diverse biological contrasts and events is displayed together, showing where and how strongly it changes. Conversely, for a given biological event, the metabolites that respond can be retrieved at once and ranked by effect. The same structure therefore supports exploration from either direction, metabolite-first or biological event/contrast-first.

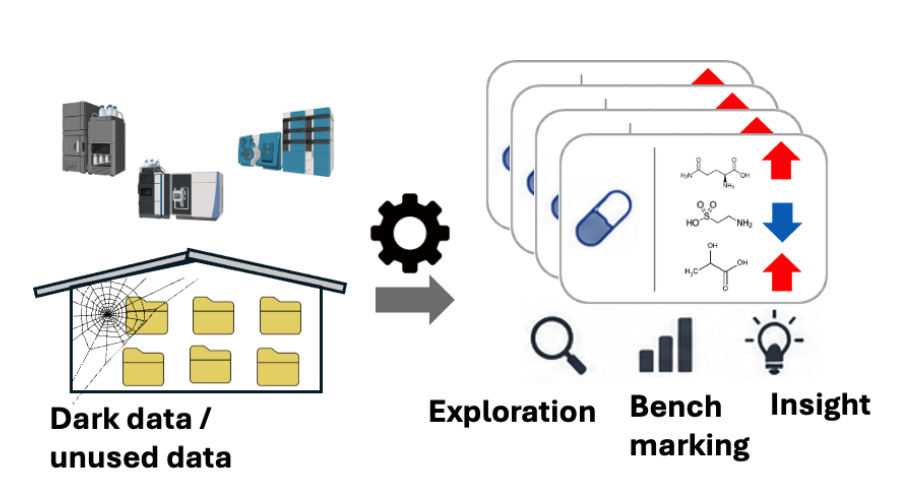


Similarity of metabolic change in each differential profile is quantified based on the iDMET method. (Matsuta R, et al, *BMC Bioinformatics*. 2022.) This similarity is then used to graph structure in which differential profiles that shows similar metabolic change are clustered. Together with extraction with keyword of interest and community detection, differential profiles with similar metabolic change can be spotted and can perform in-depth analyses like enrichment and analysis.

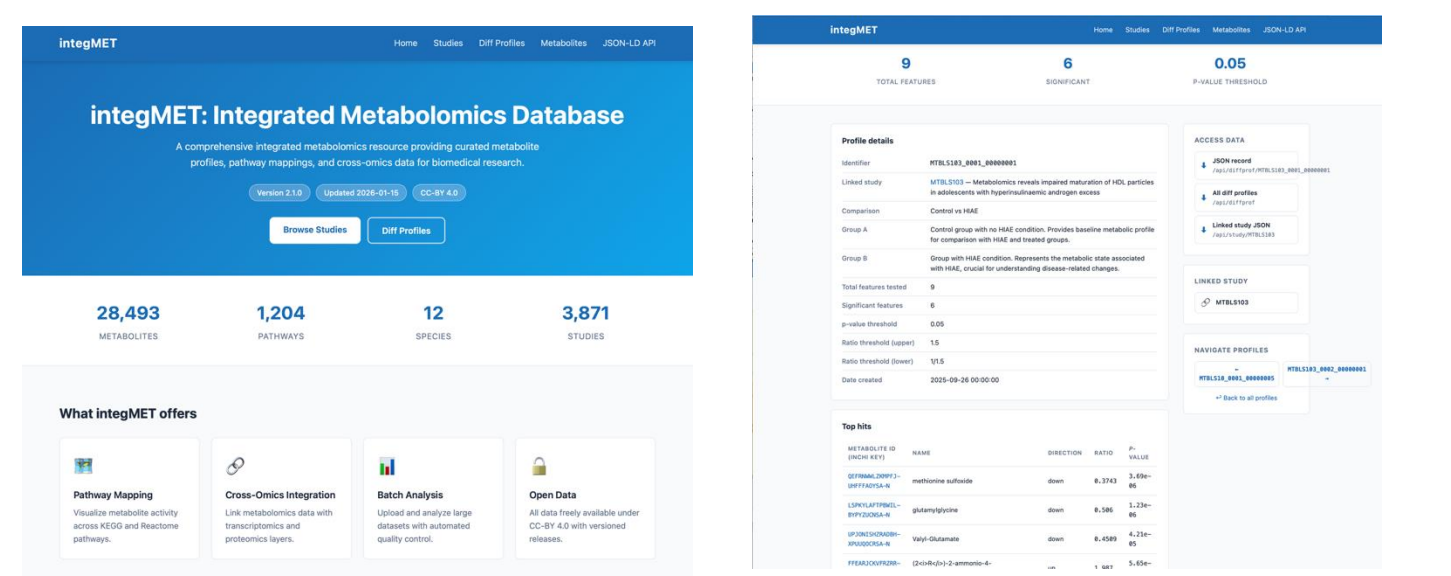
Discussion



By converting isolated, previously unusable datasets into differential profiles, this approach makes them directly comparable across the entire repository. The metabolic changes linked to a phenomenon of interest become immediately visible, and the biological events in which a given metabolic change occurs can be searched pan-repository, enabling exploratory reanalysis. The same data also support meta-analysis, pooling related contrasts to identify reliable biomarkers.



Within an organization, MS data are typically analyzed once and then shelved, rarely revisited. Applying this Differential Profile structure to internal data turns this unused data archive into a reusable asset. Teams can then ask whether a metabolic change has appeared in earlier projects, benchmark new results against accumulated history, avoid repeating experiments, and meta-analyze across their own past projects, so every dataset keeps paying off.



Landing page

Differential Profile page

The reprocessed repository was assembled into a working database organized as three linked layers — Study, Differential Profile, and Metabolite — and made navigable through a graphical interface. This differential-profile-based integrated database is currently online and publicly available at <https://integmet.org>

Acknowledgement

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



Contact

For inquiries or collaboration, on this method, online database access, or applying it to your organization's data, please contact Eisuke Hayakawa (hayakawa.eisuke531@mail.kyutech.jp)

