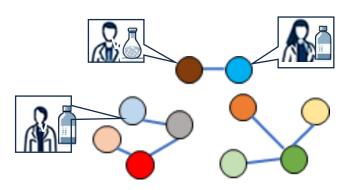
Network-based Metabolomics Data Integration: Towards Comprehensive Integrative Reanalysis

Eisuke Hayakawa

Metabolome Informatics Research Team Center for Sustainable Resource Science RIKEN

Background





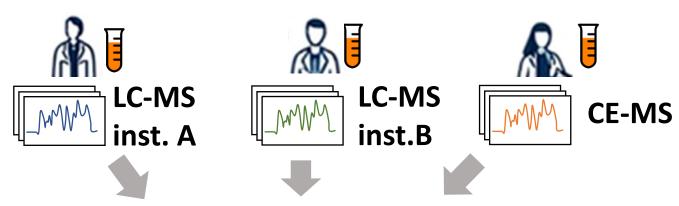
1. Expanding Scale of Metabolomics Data:

- Metabolomics has been utilized in various fields, and a vast amount of metabolome data has been accumulated so far.
- Such metabolomics data are stored in data repositories in recently, but it is not utilized beneficially.

2.Benefit of Data Integration:

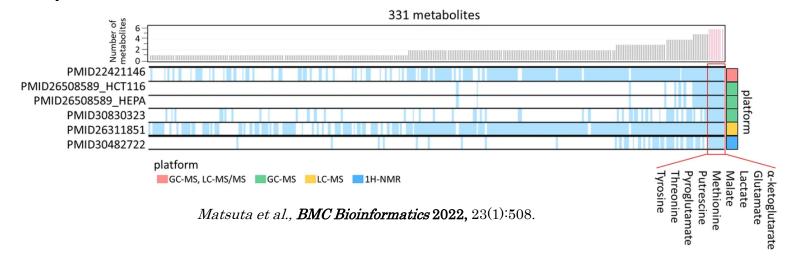
- Integrating diverse metabolomics datasets unlocks new possibilities for data reanalysis.
- Data integration enables a holistic view of metabolic processes, enhancing our understanding of complex biological mechanisms.

Technical Challenges in Data Integration

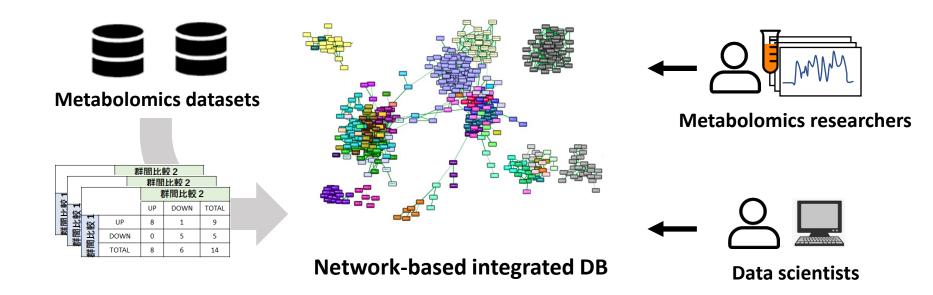


Data integration: infeasible

- Variability in instrumentation creates standardization challenges.
- Limited absolute quantification and instrument-dependent signal intensities impact data comparability.
- Metabolite detection varies with instrumental settings, affecting data consistency across platforms.



Objective: Advancing Metabolomics Data Integration



1. Metabolite differential profile-based Integration:

• Employing a differential profile-based approach to integrate metabolomics data, enabling more precise comparative analysis across different datasets.

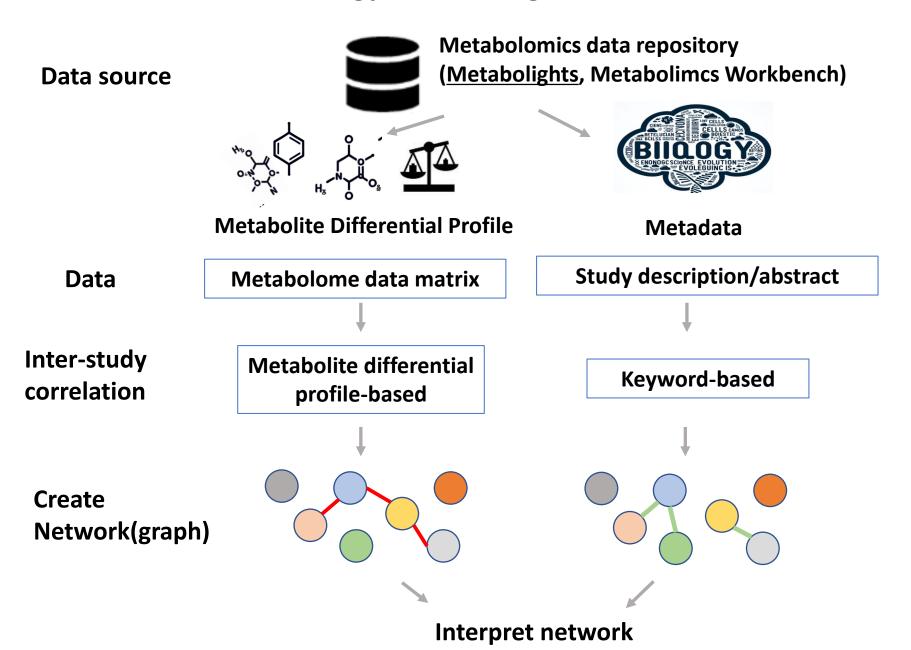
2. Metadata-Based Integration:

• Keyword-based data integration to support interpretation of complex metabolome data sets.

3. Unique Integrated Network for Data Exploration:

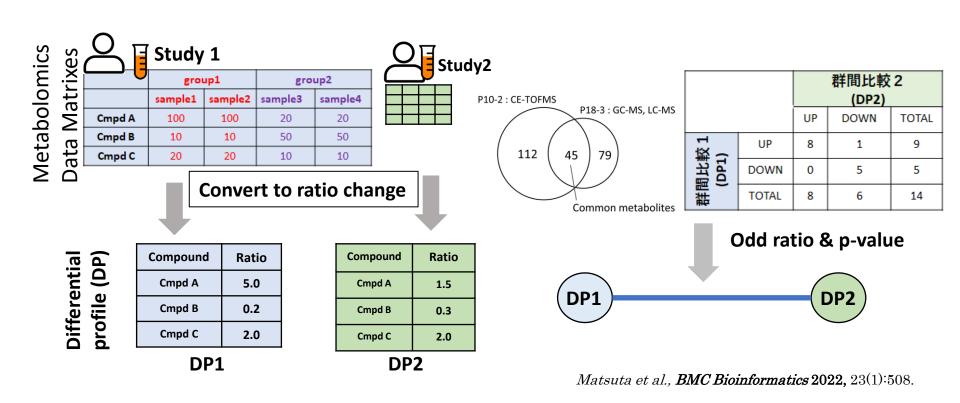
 Developing a integrated network platform that facilitates the exploration of vast metabolomics data, paving the way for novel discoveries and insights.

Strategy: Data integration



Data integration 1: Metabolome differential profile

- iDMET, a network-based approach for integrating metabolomic data from different studies by focusing on <u>differential profiles</u> between groups. (Matsuta et al *BMC Bioinformatics* 2022, 23:508)
- Creation of 2 \times 2 crosstabulation tables for pairs of differential metabolomic profiles, and Use of odds ratio for assessing correlation between pairs.



Metabolome data integration



Study 1

S_MTBLS**.txt**

Sample	Characteristic 3
C1	Control (GroupA)
C2	Control (GroupA)
N1	Drug Treated (Group B)
N2	Drug Treated (Group B)

maf.tsv

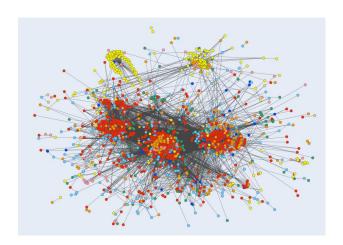
Metabolite	C1	C2	N1	N2
Estrone	1.69E+08	7.0E+08	5.7E+08	7.2E+08
Dihydrocortisol	1.52E+09	1.2E+09	3.6E+09	1.4E+09
Adipate semialdehyde	3.26E+09	1.9E+09	3.2E+09	1.4E+09
3-Aminopentanedioate	1.69E+08	7.0E+08	3.2E+09	3.6E+06
N-[(2S)-2-Amino-2-c	1.69E+08	1.4E+09	7.1E+08	3.6E+09



Create Differential Profile

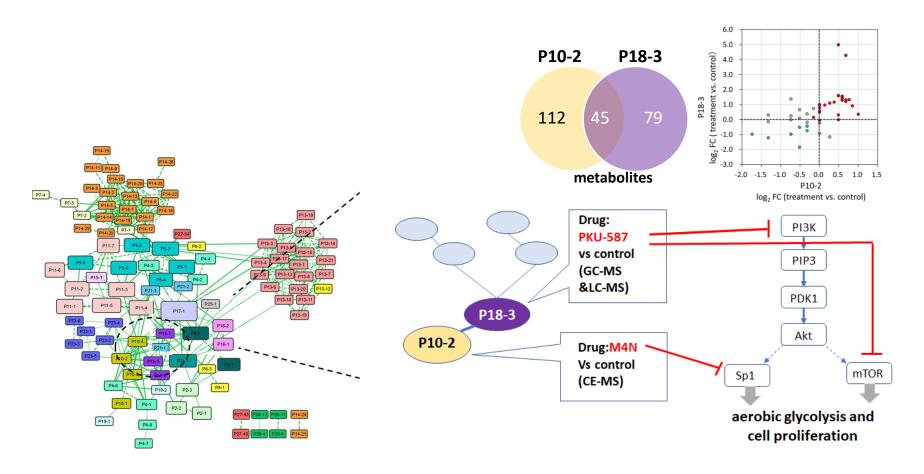






Integrated network

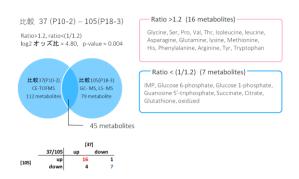
Finding overlooked correlations between studies



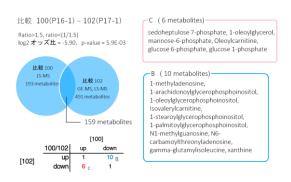
Matsuta et al., BMC Bioinformatics 2022, 23(1):508.

- iDMET detects similarity of metabolite differential profile across studies
- It enables exploration of hidden correlation of studies, leading to novel finding and new hypothesis

Other Examples demonstrating the effectiveness of data re-analysis

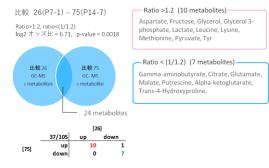


- H1975 cell +/- mTOR inhibitor
- L428 cell +/- Tmndga

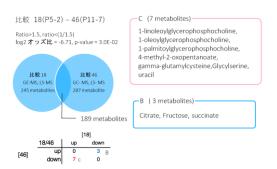


- +/- AKT inhibitor
- Liver cancer vs normal tissue

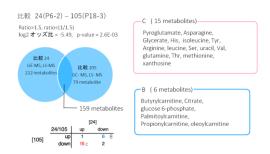
- Ratio >1.2 (16 metabolites) 比較 37 (P10-2) - 105(P18-3) Glycine, Ser, Pro, Val, Thr, Isoleucine, leucine, Ratio>1.2, ratio<(1/1.2) Asparagine, Glutamine, lysine, Methionine, log2 オッズ比 = 4.80. p-value = 0.004 His, Phenylalanine, Arginine, Tyr, Tryptophan Ratio < (1/1.2) (7 metabolites) H-8637(010-2 H-86105(018-3) IMP. Glucose 6-phosphate, Glucose 1-phosphate CE-TOFMS GC- MS, LS- MS Guanosine 5'-triphosphate. Succinate. Citrate. Glutathione, oxidized 45 metabolites
- Liver cancer vs normal tissue
- L428 cell +/1 Tmndga



- Colorectal adenocarcinoma Hypoxia vs normal
- Liver cancer Hypoxia vs normal



- +/- Rep Sen
- Leukemia +/- radiation



- ccRCC cell vs control
- MDA-MD-231cell hypoxia vs normal

Detecting common metabolic alterations across various cancer-related studies.

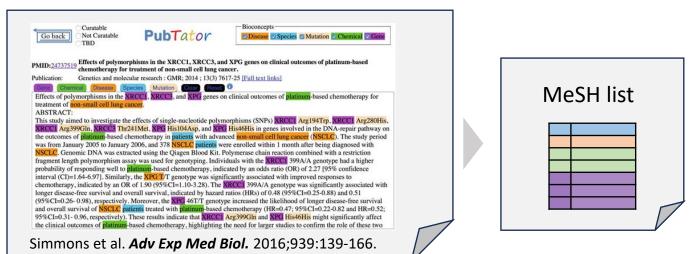
Data integration 2 : concept/keyword-based

- Metabolomics data repository hosts diverse studies, each with unique backgrounds. Difficult to find similarities and connections between studies.
- We use tools to annotate key biological concepts (diseases, species, genes and chemicals) in studies.
- Annotated concepts are used to assess and quantify correlations between studies.

Pubtator

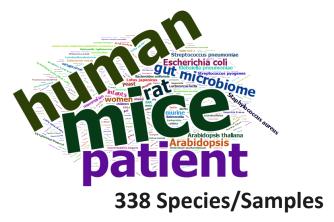
- Tool designed for the automated annotation of biomedical literature.
- It uses text-mining technologies to identify and annotate key biological entities (concepts) within abstracts and full-text articles

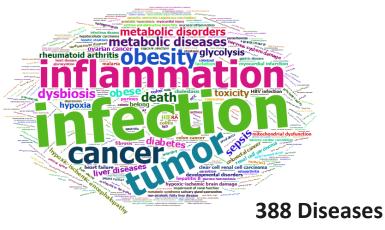
Wei et *Nucleic Acids Research* 47, 2019

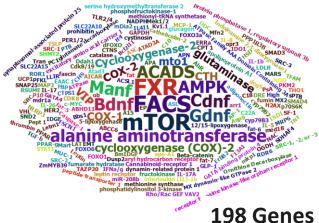


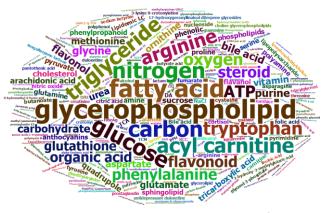
Data Mining Metabolome Data repository











696 Chemicals

4 categories of keywords were extracted from study description of Metabolights

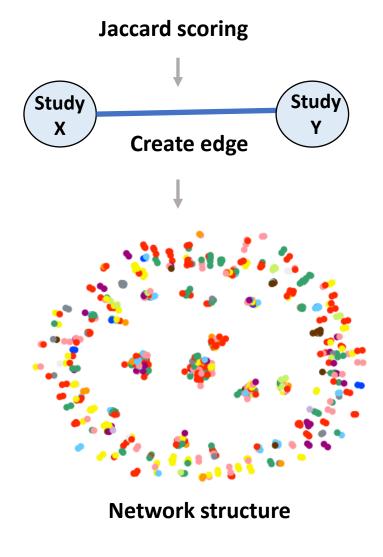
Data integration: Metadata-based



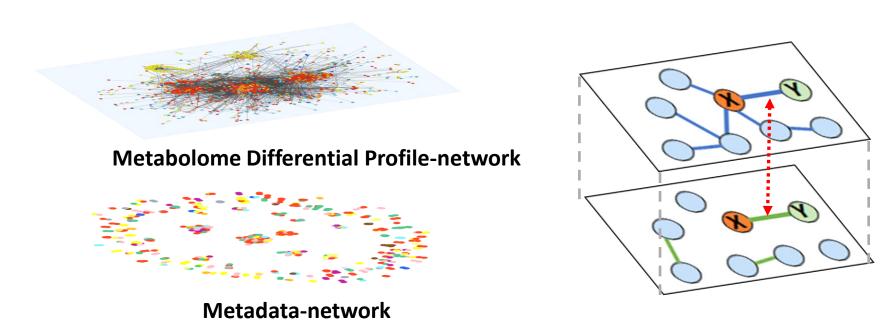
Study X

Carcinoma cells derived from Human were treated with Drug X together with Drug Y. The treated cells were exposed to condition M and their growth rate were recorded. We also observed expression level of genes (gene L, gene M and gene N).

Pubtator Study X MeSH list



Two distinct network structures

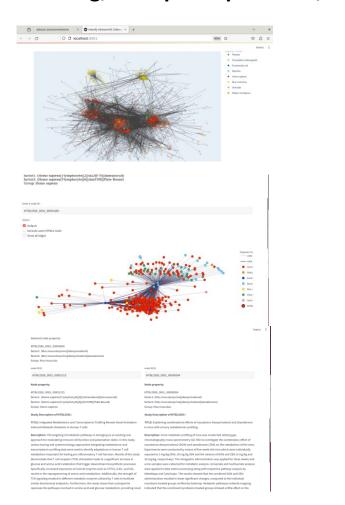


Combined Insight from Metabolic Changes and Concepts:

- •The metabolic change network provides insights into how different studies relate in terms of actual metabolic changes.
- •The metadata-based network offers a perspective on conceptual and thematic similarities across studies.
- •While metabolic changes provide direct biological insights, concept similarities can reveal broader thematic connections, potentially providing new hypotheses or overlooked relationships.

Development as a tool/database

- It's essential to provide open access to integrated datasets for comprehensive analysis by the scientific community.
- Developing a dedicated interface is crucial for facilitating enhanced understanding, in-depth exploration, and effective visualization of data.





Code in general



App framework

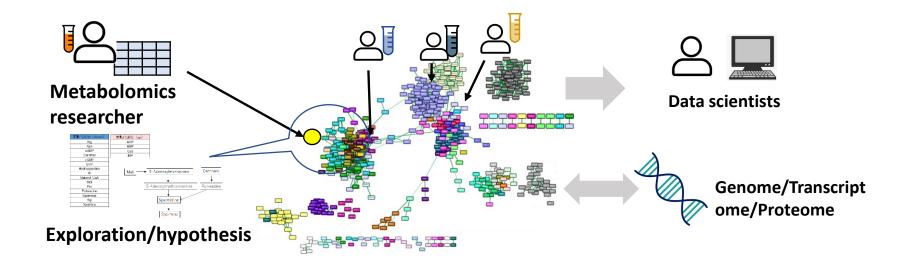


Graph DB



Visualization

Summary



- Integrate large metabolome datasets to enables a holistic view of metabolic processes, enhancing our understanding of complex biological phenomena.
- This unique network-based platform significantly contributes to the metabolomics/science community by simplifying and enhancing accessibility for data reanalysis

Acknowledgements

Ms. Mikiko Takahashi

Mr. Yutaka Yamada

Prof. Masanori Arita

(Metabolome Informatics Research Team)



Dr. Hiroyuki Yamamoto (Human Metabolome Technologies)

Ms. Rita Matsuta (Human Metabolome Technologies)



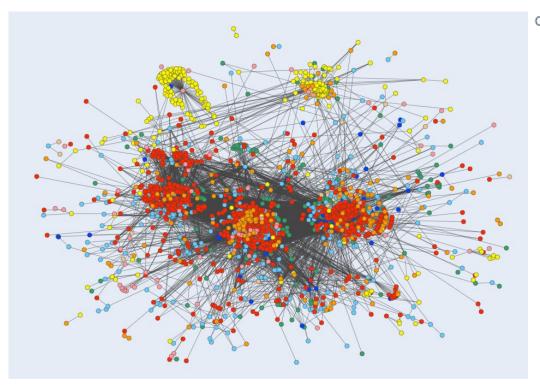
Mr. Kozo Nishida (Tokyo University of Agriculture and Technology)

Prof. Hiroshi Tsugawa (Tokyo University of Agriculture and Technology)



Prof. Shin Kawano (Kitasato university)





Organism Group

- Plantae
- Drosophila melanogaster
- Escherichia coli
- Bacteria
- Homo sapiens
- Mus musculus
- Animalia
- Rattus norvegicus

Metabolights stat

Taxon_curated	個数 / Taxon_curated]
9606	40	8 Homo sapiens
10090	20	5 Mus musculus
Mixture	6	7 Mixture
3702	4	9 Arabidopsis thaliana
10116	3	9 Rattus norvegicus
environment	3	4 environment
9913	2	9 Bos taurus
29760	2	7 Vitis vinifera
562	2	4 Escherichia coli
Other	2	3 Other
4932	1	9 Saccharomyces cerevisiae
4081	1	8 Solanum lycopersicum
9823	1	7 Sus scrofa
7227	1	3 Drosophila melanogaster
9940	1	2 Ovis aries
4522	1	1 Lolium perenne
4530	1	1Oryza sativa
5691	1	0 Trypanosoma brucei
6239		9 Caenorhabditis elegans