EnrichMET: R Package for Integrated Pathway and Network Analysis for Metabolomics

Yonatan Ayalew Mekonnen[1] (yonatanayalew.mekonnen@moffitt.org), Neha Dhake[2](nehadhake@usf.edu), Shreya Jaiswal[3](jais2157@stthomas.edu), Vanessa Rubio[1] (Vanessa.Rubio@moffitt.org), Isis Narváez-Bandera[1] (Isis.Narvaez-Bandera@moffitt.org), Hayley Ackerman[1,4] (Hayley.Ackerman@moffitt.org), Elsa Flores[1,4] (Elsa.Flores@moffitt.org), Paul Stewart[5,6 *] (Paul.Stewart@hci.utah.edu)

[1] Department of Molecular Oncology, Moffitt Cancer Center, Tampa, Florida, USA; [2]Department of Bioinformatics, University of South Florida, Tampa, Florida, USA; [3]University of St. Thomas, Department of Computer Science, Minnesota, USA; [4]Cancer Biology and Evolution Program, Moffitt Cancer Center, Tampa, Florida, USA; [5]Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City UT, USA. [6]Huntsman Cancer Institute, University of Utah, Salt Lake City UT, USA.

*Corresponding Author

Abstract

Advances in metabolomics have significantly improved our understanding of cellular processes by enabling the identification of hundreds of metabolites in a single experiment. These developments provide valuable insights into complex metabolic networks. Pathway Enrichment Analysis (PEA) is essential for connecting identified metabolites with biological pathways, uncovering meaningful relationships, and revealing underlying metabolic mechanisms. However, existing tools, often require multiple steps, rely on web-based interfaces, and require R packages that can be challenging to install due to outdated or missing dependencies. This hampers both reproducibility and efficiency. To overcome these limitations, we introduce EnrichMet, an R package designed for efficient and reproducible pathway, metabolite set enrichment analysis (MetSEA). The package integrates fgsea for rapid MetSEA and igraph for topology-based metrics, enabling comprehensive integration of metabolomics data. EnrichMet utilizes curated pathway data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and employs Fisher's Exact Test to identify significantly enriched pathways. Through a case study on experimental metabolomics datasets, we demonstrate the robustness and efficiency of EnrichMet, benchmarking it against existing tools. The results show that EnrichMet provides accurate and comprehensive pathway enrichment outcomes while reducing computational time and user effort. Moreover, its flexible framework allows EnrichMet to be adaptable to a broad range of data types and applications beyond KEGG. By streamlining the pathway analysis process, EnrichMet enables researchers to derive deeper insights from metabolomics data, promoting reproducible and efficient research practices.

Availability and implementation: (https://github.com/biodatalab/enrichmet.git)

Introduction

High-throughput metabolite identification in a single experiment reveals unprecedented insights into the complexity of cellular processes. However, interpreting these metabolites in a meaningful biological context remains challenging[1]. Pathway Enrichment Analysis (PEA)[2, 3] and Metabolite Set Enrichment Analysis[4] address this challenge by mapping metabolites to known biochemical pathways. These methods statistically evaluate whether certain pathways contain more observed metabolites than would be expected by chance, thereby highlighting metabolic processes that may be altered or dysregulated processes[5]. Pathway enrichment transforms complex, high-dimensional metabolomics data into concise, biologically relevant insights. Building on this concept, our tool provides an intuitive and streamlined approach to enrichment analysis, which helps users to uncover

functional patterns and drive meaningful insights from their metabolomics data with greater confidence. These approaches not only streamline data interpretation but also highlight potential biomarkers and therapeutic targets in various biological studies. Tools such as MetaboAnalyst[6], Mummichog[7], PIUMet[8], PUMA[9], and FELLA[10] exist for performing pathway enrichment analysis, with MetaboAnalyst being among the most widely used platforms[11]. However, installing the package can be challenging due to outdated or missing dependencies. Moreover, while well-suited for advanced users, the package provides limited built-in options for visualizing analysis outputs. A comparison of existing tools is summarized in Table 2.

EnrichMet: integrated pathway and network analysis

While comprehensive, existing tools often involve multiple steps and web-based interactions, which incur substantial run-time and limit reproducibility. Moreover, these tools may require prior expertise in statistical and computational methods, creating barriers for researchers new to the field. To overcome these challenges, we developed EnrichMet, an R package designed for fast, reproducible, user-friendly pathway, MetSEA and topological analyses. EnrichMet leverages curated pathway data from the KEGG and applying Fisher's Exact Test for overrepresentation analysis and utilizes the fgsea algorithm to perform MetSEA efficiently. By integrating metabolite input and background datasets, EnrichMet identifies enriched pathways with minimal user intervention, requiring only a single function call. The full workflow is illustrated in Fig. 1. This manuscript highlights EnrichMet's streamlined design, statistical rigor, and adaptability, demonstrating its utility for metabolomics data analysis through a case study dataset. Furthermore, we compare its performance to MetaboAnalystR and FELLA, showcasing its efficiency and ease of use for researchers in diverse scientific domains.

Pathway enrichment analysis functionality

Metabolite and pathway data were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, specifically targeting on human-specific pathways to ensure biological relevance. The analysis utilized two primary datasets: an input list of differentially expressed metabolites identified in experimental studies and a comprehensive background set containing all known KEGG metabolites associated with human KEGG pathways, independent of experimental observations. These datasets were organized to map each pathway to its corresponding metabolites, providing a robust framework for statistical analysis. For each pathway, a 2x2 contingency table was constructed as shown in Table1. Fisher's Exact Test [12] was applied to the contingency table to assess whether the number of identified metabolites in a pathway significantly exceeded expectations. The output of Fisher's Exact Test is a p-value, which indicates the likelihood that the observed enrichment or depletion of metabolites in a pathway is due to random chance. To account for multiple comparisons and enhance the reliability of statistical significance, p-values were adjusted using the Benjamini-Hochberg method and Q-values were also computed to provide an additional measure of significance. The final output ranks pathways by significance, highlighting those enriched with the identified metabolites.

The p value from Fisher's Exact Test is calculated as:

$$P = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,n!}$$

Where, a, b, c, d are counts from the contingency table.

n= a+b+c+d is the total number of metabolites

Results from the enrichment analysis were summarized in tabular format and visualized using dot plots of -log10(adjusted p-values), highlighting the top significantly enriched pathways. To further

assess its performance, results were compared against those generated using MetaboAnalystR and FELLA R packages, demonstrating consistency in findings while showcasing a more streamlined and reproducible analytical process.

Metabolite set enrichment analysis (MetSEA) functionality

Pathway enrichment analysis typically relies on predefined significance thresholds to select metabolites or feature, which can be arbitrary and introduce bias [13]. In contrast MetSEA leverages the entire ranked list of metabolites based on the fold change and p-values, without requiring the hard cutoffs. This approach enables the detection of coordinated and subtle changes across pathways, offering a more systematic and less biased assessment of biological enrichment[4]. EnrichMet also performs metabolite set enrichment analysis (MetSEA) by integrating the fgsea R package. Pathway data is formatted into a GMT file format to ensure compatibility. Metabolites are ranked based on their differential expression metrics (log2 fold change and p-values), creating a ranked list required for fgsea. After ranking, the running sum is calculated as follows:

If metabolite p is not in set S, then the score is decreased by,

$$X_i = \frac{1}{N - N_s}$$

and if metabolite p is in set S, then the score is increased by:

$X_i = \frac{|r_i|^p}{N_R}$ Where, $N_R = \sum_{i \in S} |r_i|^p$ and, r_i represent the ranking metric for metabolite i.

Enrichment Score (EC) represents the maximum deviation from zero observed as the algorithm walks down the ranked list of the metabolites. The normalized enrichment score (NES) is then calculated by dividing the ES by the mean of ES values obtained from permutation, thereby accounting for differences in the pathway size and variability[14]. The analysis outputs significant pathways, their adjusted p-values, and enrichment statistics. These results are then exported to Excel files, providing a comprehensive summary of enriched pathways and their associated metrics.

Topological measure functionality

To measure metabolite centrality within biological network, we employ the igraph package in R to computes graph-based centrality metrics. Betweenness Centrality (BC) measures the number of shortest paths from all node pairs that pass through a given node, highlighting its role in facilitating communication within the network. Hub metabolites are identified based on centrality, specifically the number of edges connected to a metabolite to others within the network. The metabolite network is represented as a graph, where nodes correspond to metabolites and edges denote shared pathway associations, thereby capturing the topological relationships among metabolites. Betweenness centrality is calculated as:

$$BC(v) = \sum \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$RBC(v) = \frac{BC(v)}{(\mathrm{N}-1)(\mathrm{N}-2)/2}$$

where σ_{st} is the total number of shortest paths between nodes s and t, $\sigma_{st}(v)$ is the number of those paths that pass through node v. To normalize BC, we compute relative betweenness centrality (RBC) which is normalized form of BC and N is the total number of nodes in the network. To account the relative importance of pathways in a global metabolic network, pathway impact score was calculated as:

$$Pathway\ impact = \frac{\sum RBC\ of\ matched\ metabolites}{\sum RBC\ of\ all\ metabolites\ in\ the\ pathway}$$

Additionally, we construct a metabolite network based on the known chemical-chemical relationships from STITCH database reflecting biological meaningful associations between metabolites.

Case study of KrasG12D mutation

In this paper, we used summary statistics of the differential analysis derived from metabolomic measurements of KrasG12D mutation and KrasG12D mutation with TAp73 deletion. We selected the top 50 metabolites to apply Fisher's Exact Test for pathway enrichment analysis. After adjusting for multiple comparisons using the Benjamini-Hochberg method, pathways with an adjusted p-value < 0.05 were considered significantly enriched. The dot plot illustrates the -log10(adjusted p-values) of the most enriched pathways, highlighting key biological processes in Fig. 2a, 3a and 3b. The pathway impact plot shows prioritizing pathways that are not only statistically enriched but also structurally critical in the biological networks (Fig2. b). Therefore, the pathway enrichment and impact analysis in alignment with the MetSEA results (Fig. 2c), identifies three main categories: metabolic reprogramming and energy production, nutrient uptake and homeostasis, and hormone and xenobiotic processing. The metabolites involved in these pathways are visualized in Figures 3a and 3b. These categories are closely associated with KrasG12D-driven tumors and the metabolic reprogramming resulting from TAp73 deletion[15]. The loss of TAp73 disrupts processes like mineral absorption, protein digestion, and adipocyte lipolysis [16, 17]. Additionally, Hormonal regulation and xenobiotic metabolism also influence stress responses, drug metabolism, and systemic homeostasis[18].

These findings highlight potential therapeutic strategies for KrasG12D/TAp73-deficient lung cancer, focusing on targeting lipid metabolism, nucleotide biosynthesis, and drug resistance mechanisms. Notably, key hub metabolites identified in the network, such as glucose-6-phosphate (Fig. 4a, b), Arachidonic acid (Fig. 4a, b, c), and L-serin (Fig 4a, b) are associated with tumor growth and may serve as promising therapeutic targets [19-21]. Additionally, L-phenylalanine, palmitic acid, and adrenic acid have been implicated in tumor progression through various mechanisms, including metabolic reprogramming, membrane biosynthesis, and modulation of inflammation, underscoring their roles as potential metabolic vulnerabilities [22-24]. Furthermore, the chemical-chemical interaction plot (Fig. 4c) identifies Xanthine as a contributor to the unfolded protein response and autophagic degradation, supporting the survival of cancer cell under nutrient-deprived conditions[25].

Benchmarking

Streamlined Workflow: MetaboAnalyst is widely regarded as the gold standard for metabolomics and in many aspects of data analysis in metabolomics. However, installing the package can be challenging due to outdated or missing dependencies. Similarly, the FELLA package requires building a local KEGG graph database, which is memory intensive and time consuming. This can significantly slow down performance, making FELLA less practical for high-throughput metabolomics datasets. However, EnrichMet offers a streamlined alternative for pathway enrichment analysis by performing the entire process with a single R function call. This one-step workflow reduces manual steps such as data formatting, and parameter selection minimizing user intervention, reducing potential errors and improving reproducibility.

Speed and Efficiency: EnrichMet significantly outperforms MetaboAnalystR and FELLA in processing time. For representative dataset of 50 metabolites, EnrichMet completed the pathway analysis in

2.71 seconds, compared to 27.16 seconds of MetaboAnalystR and 13.24 minutes for FELLA. This substantial performance advantage makes EnrichMet particularly ideal for high-throughput or iterative analyses, where speed is critical.

Transparency: EnrichMet offers full transparency and control over the analysis process, allowing users to easily access results and integrate the package into larger R workflows.

Reproducibility: EnrichMet runs entirely within the R environment, providing full control over reproducibility and customization. This is critical features for modern data-driven research.

Summary

By reducing the number of steps required for pathway enrichment analysis, EnrichMet enhances the user experience, enabling scientists to focus on interpreting results rather than navigating complex interfaces. This ease of use, combined with robust statistical methods, makes EnrichMet an invaluable tool for metabolomics research and pathway analysis.

In conclusion, by integrating pathway enrichment analysis, topological measures, and MetSEA, EnrichMet offers a comprehensive and robust analysis, providing reliable results for comparison.

Reference

1. Johnson, C.H., J. Ivanisevic, and G. Siuzdak, *Metabolomics: beyond biomarkers and towards mechanisms.* Nat Rev Mol Cell Biol, 2016. **17**(7): p. 451-9.
2. Wieder, C., et al., *Pathway analysis in metabolomics: Recommendations for the use of over-representation analysis.* Plos Computational Biology, 2021. **17**(9).
3. Lu, Y., Z. Pang, and J. Xia, *Comprehensive investigation of pathway enrichment methods for functional interpretation of LC-MS global metabolomics data.* Brief Bioinform, 2023. **24**(1).
4. Xia, J. and D.S. Wishart, *MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W71-7.
5. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.
6. Pang, Z., et al., *MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics.* Nat Commun, 2024. **15**(1): p. 3675.
7. Li, S., et al., *Predicting network activity from high throughput metabolomics.* PLoS Comput Biol, 2013. **9**(7): p. e1003123.
8. Pirhaji, L., et al., *Revealing disease-associated pathways by network integration of untargeted metabolomics.* Nat Methods, 2016. **13**(9): p. 770-6.
9. Hosseini, R., et al., *Pathway-Activity Likelihood Analysis and Metabolite Annotation for Untargeted Metabolomics Using Probabilistic Modeling.* Metabolites, 2020. **10**(5).
10. Picart-Armada, S., et al., *FELLA: an R package to enrich metabolomics data.* BMC Bioinformatics, 2018. **19**(1): p. 538.
11. Chong, J., D.S. Wishart, and J. Xia, *Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis.* Curr Protoc Bioinformatics, 2019. **68**(1): p. e86.
12. Fisher, R.A., *The design of experiments*. 1935, Edinburgh, London,: Oliver and Boyd. xi, 252 p.
13. Tsouka, S. and M. Masoodi, *Metabolic Pathway Analysis: Advantages and Pitfalls for the Functional Interpretation of Metabolomics and Lipidomics Data.* Biomolecules, 2023. **13**(2).
14. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
15. Srivastava, S., et al., *The Making of a PreCancer Atlas: Promises, Challenges, and Opportunities.* Trends Cancer, 2018. **4**(8): p. 523-536.
16. Gonfloni, S., V. Caputo, and V. Iannizzotto, *P63 in health and cancer.* Int J Dev Biol, 2015. **59**(1-3): p. 87-93.
17. Rufini, A., et al., *TAp73 depletion accelerates aging through metabolic dysregulation.* Genes Dev, 2012. **26**(18): p. 2009-14.
18. Brunton, L.L. and B.r.C. Knollmann, *Goodman & Gilman's the pharmacological basis of therapeutics*. Fourteenth edition. ed. 2023, New York: McGraw Hill. xviii, 1645 pages.
19. Lan, T., et al., *Glucose-6-phosphate dehydrogenase maintains redox homeostasis and biosynthesis in LKB1-deficient KRAS-driven lung cancer.* Nat Commun, 2024. **15**(1): p. 5857.
20. Wang, D. and R.N. Dubois, *Eicosanoids and cancer.* Nat Rev Cancer, 2010. **10**(3): p. 181-93.
21. Tajan, M., et al., *Serine synthesis pathway inhibition cooperates with dietary serine and glycine limitation for cancer therapy.* Nat Commun, 2021. **12**(1): p. 366.
22. Liang, K.H., et al., *Plasma phenylalanine and glutamine concentrations correlate with subsequent hepatocellular carcinoma occurrence in liver cirrhosis patients: an exploratory study.* Sci Rep, 2020. **10**(1): p. 10926.
23. Wang, X., C. Zhang, and N. Bao, *Molecular mechanism of palmitic acid and its derivatives in tumor progression.* Front Oncol, 2023. **13**: p. 1224125.

24. Kagan, V.E., et al., *Oxidized arachidonic and adrenic PEs navigate cells to ferroptosis.* Nat Chem Biol, 2017. **13**(1): p. 81-90.
25. Chen, M.M., et al., *Xanthine dehydrogenase rewires metabolism and the survival of nutrient deprived lung adenocarcinoma cells by facilitating UPR and autophagic degradation.* Int J Biol Sci, 2023. **19**(3): p. 772-788.

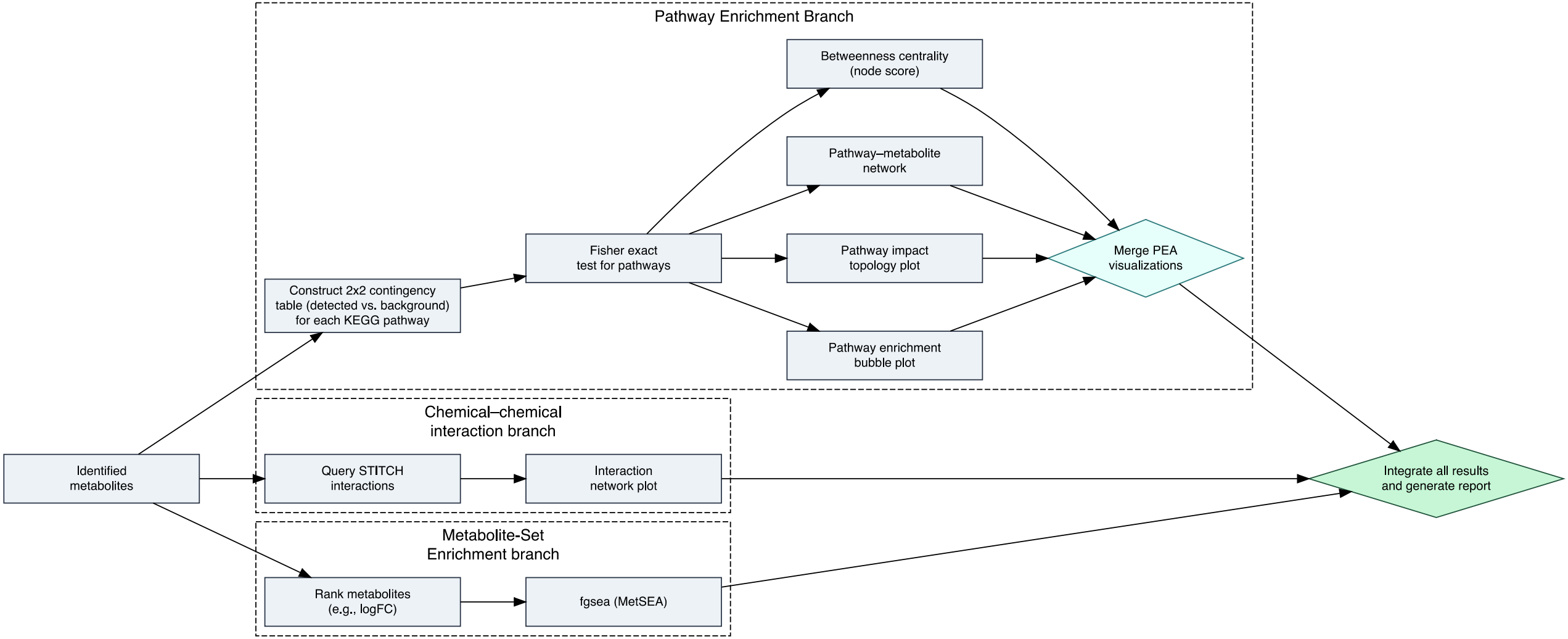

Figure 1. Overview of Enrichmet data analysis workflow. The workflow comprised three main branches: pathway enrichment, chemical-chemical interaction and metabolite set enrichment branch. The pathway enrichment branch builds 2×2 contingency tables (detected vs. background metabolites) for each KEGG pathway, followed by Fisher's exact test to identifysignificantly enriched pathways. Concurrently, pathway enrichment bubble plots, pathway impact topology plots, pathway–metabolite network graphs, and betweenness centrality plots are generated to highlight key nodes. The chemical–chemical interaction branch queries STITCH for metabolite interaction networks, producing interaction network plots. In parallel, ranked metabolite features (e.g., log fold changes) are analyzed using Metabolite Set Enrichment Analysis (MetSEA) with the fgsea algorithm. Finally, all results are integrated into a comprehensive report.

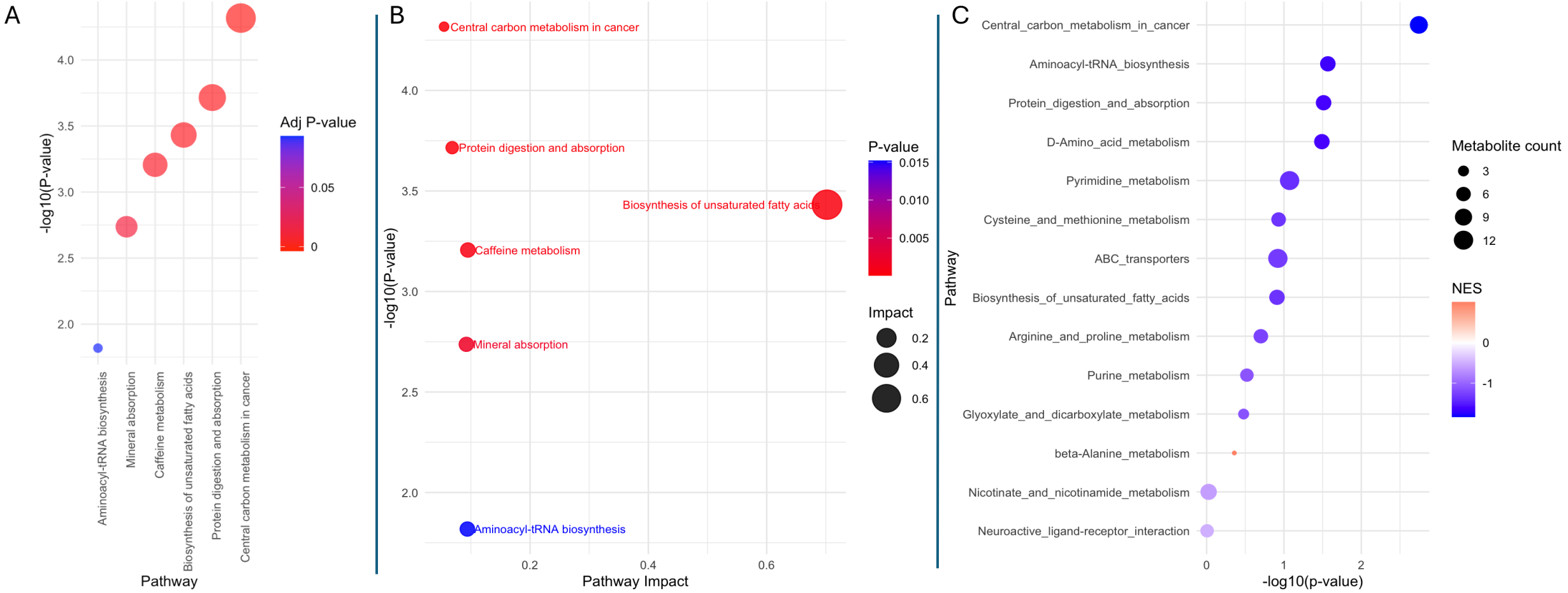

Figure 2. Overview of pathway enrichment and topology-based analysis results.
A) Pathway enrichment analysis displaying pathways ranked by $-\log_{10}$(p-value), calculated using Fisher's exact test based on input metabolite-to-pathway mappings.
B) Pathway impact plot integrating enrichment significance with topological information, where Relative Betweenness Centrality (RBC) highlights the most topologically influential pathways.
C) Metabolite Set Enrichment Analysis (MetSEA), showing significantly enriched metabolite sets derived from ranked metabolite data.

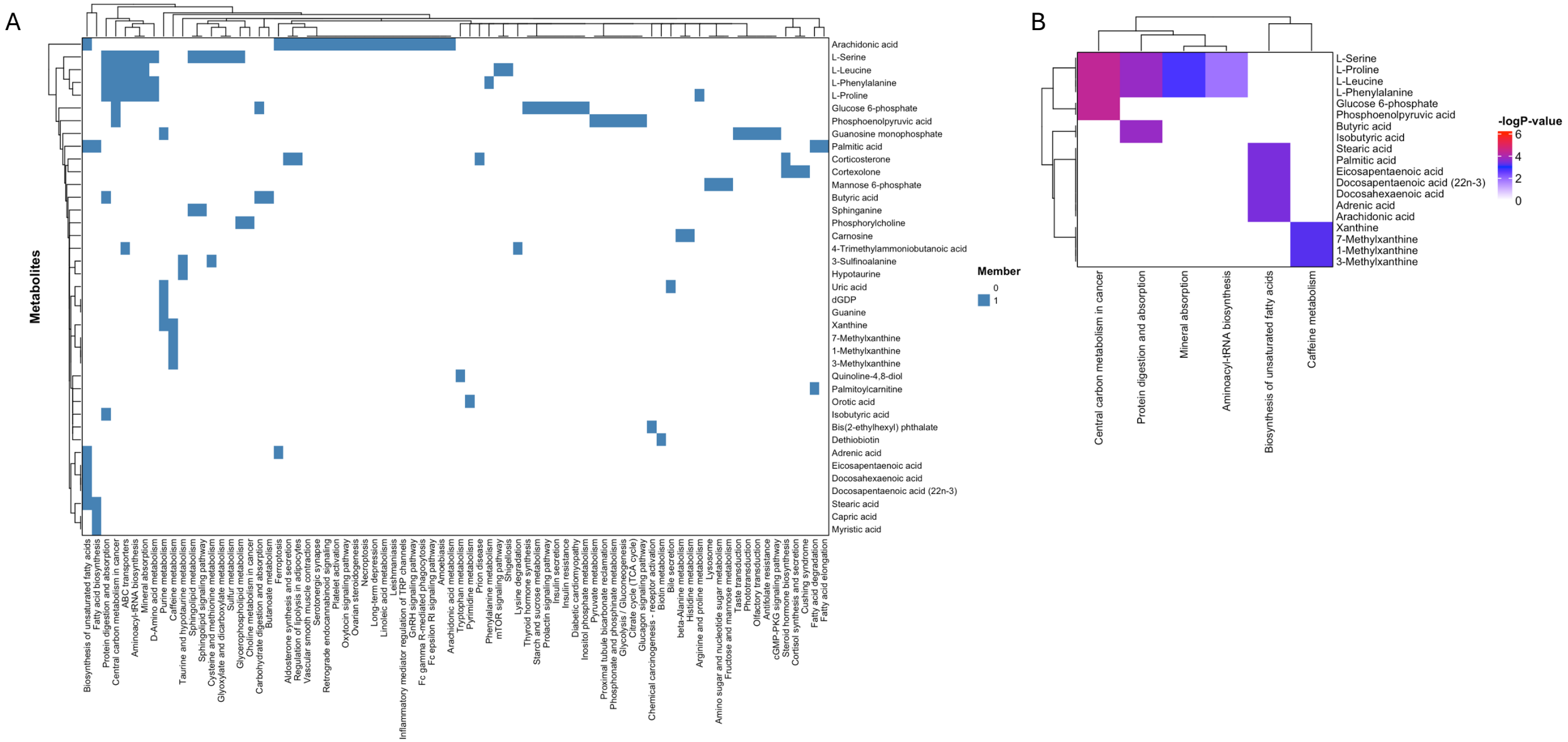

Figure 3. Visualization of pathway enrichment results and metabolite associations.

A) Pathway-metabolite association plot illustrating each pathway alongside its corresponding mapped metabolites.

B) Heatmap showing a gradient of $-\log_{10}$(p-values), representing the significance of enriched pathways. Only pathways with at least two mapped metabolites are included

Figure 4. Network-based visualization of metabolite importance and interactions.
A) Relative betweenness centrality of input metabolites, reflecting their topological importance and potential regulatory roleswithin the metabolic network.
B) Metabolite-pathway interaction network, illustrating the connectivity between pathways and input metabolites and highlighting their functional associations. The size of the hub metabolites is corresponded to the value of RBC, edges are pathways and nodes are metabolites.
C) Chemical-chemical interaction network derived from STITCH. Metabolites are shown as nodes whose size scales with degree (the number of direct interaction partners). Edges represent experimentally validated or predicted biochemical relationship; and edge opacity is mapped to the rescaled STITCH combined score (weight), with darker edges indicating higher-confidence interaction.

Table 1. Contingency table used for Fisher exact test in pathway enrichment analysis

	In Pathway	**Not in Pathway**	**Total**
Input Metabolite Set	a	b	a+b
Background set	c	d	c+d
total	a+c	b+d	a+b+c+d

a: identified metabolites in the pathway
b: identified metabolites not in the pathway
c: background metabolites in the pathway (not identified)
d: background metabolites not in the pathway
n: a+b+c+d is the total number of metabolites

Table 2. Comparative overview of software tools used in metabolomics pathway enrichment analysis.

Tools	Methods	Input data	Pathway Topology	Visualization Options	R package	Shiny App	Web Interface	Fast for Large Data	Independent of DB Building	Yes Score
EnrichMet	Fisher's exact test, fgsea	Identified metabolites	yes	yes	yes	Yes	no	yes	yes	86%
MetaboAnalystR	Fishers' exact test, hypergeometric test, GSEA,	Identified metabolites	yes	Moderate	yes	no	yes	moderate	yes	57%
Mummichog	Permutation-based enrichment	m/z values	yes	limited	no	no	no	yes	yes	43%
FELLA	Diffusion over KEGG graph, hypergeometric test	Identified metabolites	yes	limited	yes	no	no	no	no	29%
PIUMet	Bayesian graphical model	m/z, RT	yes	no	no	no	yes	no	no	29%
PUMA	Network-based scoring	Metabolite, gene, protein data	yes	no	no	no	no	no	no	14%