nature portfolio

Peer Review File

TidyMass2: Advancing LC-MS Untargeted Metabolomics Through Metabolite Origin Inference and Metabolic Feature-based Functional Module Analysis

Corresponding Author: Professor Xiaotao Shen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have reviewed the manuscript on TidyMass2. Compared with the first-generation TidyMass, the manuscript clearly describes the advanced functionalities of the updated framework, which makes the logic of the paper straightforward to evaluate. The authors highlight three main contributions: (1) the construction and utility of MetOriginDB, (2) an MS2-free module analysis pipeline, and (3) a user-friendly pipeline through Shiny apps. Among these, I focused my evaluation on points (1) and (2). Below, I provide comments and concerns regarding the manuscript.

1. Accessibility of MetOriginDB
The necessity of MetOriginDB is well understood, and this resource would be highly valuable to the broader community. However, it appears that the database is currently only provided as an R binary format (.rda). For third-party evaluation, it is essential that the authors also provide an ASCII-accessible format (e.g., Excel or CSV or TSV). Without such accessibility, independent validation of this resource is not feasible.

2. Resolution of origin information
For researchers studying the microbiome, origin information at the class or species level would be highly desirable. In this regard, MicrobeMASST currently appears to offer more granular classification. The authors should more clearly discuss the strengths and limitations of their approach in comparison. It is, of course, also true that metabolome information at the class or species level is not yet systematically available in public databases. Therefore, updating the discussion with the latest developments in microbiome metabolomics databases, explaining why such information cannot yet be incorporated, and describing the future directions of MetOriginDB would strengthen the manuscript.

3. Peak picking and alignment in TidyMass
It is not clear whether TidyMass (and TidyMass2) rely on the XCMS package for peak picking and alignment of raw MS data. This might have been described in the first-generation paper, but I recommend that the authors clarify this explicitly. If the algorithms were implemented from scratch, that should also be clearly stated.

4. Similarity to MetDNA
The idea of the module-based functional analysis strongly resembles the previously reported MetDNA framework by the corresponding author, Xiaotao Shen. While there may be updated features in detection or processing, the conceptual foundation appears to be the same. It would be appropriate to cite the earlier study explicitly and emphasize what specific improvements TidyMass2 offers beyond MetDNA.

5. Differentiation from existing pipelines
While the utility of the TidyMass2 pipeline is clear, it remains somewhat unclear how its ideas and functionalities fundamentally differentiate it from existing pipelines. Experienced bioinformatics researchers might already be able to produce similar outputs without using TidyMass2. Explicitly articulating the unique aspects of TidyMass2 in terms of innovation and differentiation would make the contribution clearer.

6. Validation on public datasets

If the authors wish to emphasize the utility of TidyMass2 (which I greatly respect), I recommend validating the pipeline not only on the in-house urine metabolomics dataset from human pregnancy but also on public datasets stored in repositories such as Metabolomics Workbench or MetaboLights. Since the current dataset originates from the authors' group, they are very familiar with its properties, which may bias the demonstration. In contrast, analyzing external datasets would allow the community to better assess whether the pipeline is truly "globally applicable." For example:

(A) Can the pipeline robustly handle diverse raw MS datasets without errors, as MS-DIAL and MZmine have demonstrated?
(B) Through re-analysis of public datasets, to what extent can TidyMass2 extract new biological insights that were not reported in previous studies?

Including two or three such additional validation examples would greatly strengthen the manuscript and convincingly demonstrate the broader applicability of TidyMass2.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
he article: "TidyMass2: Advancing LC-MS Untargeted Metabolomics Through Metabolite Origin Inference and Metabolic Feature-based Functional Module Analysis" by Wang and coworkers, presents an original soft-ware package for metabolic feature annotation and determination of analyte origin.
The authors describe a rise in "interpretable" features from 5.8% to 58.8%, achieved by abandoning MS^2-based annotations from matched fragmentation spectra in favor of MS^1 accurate m/z matching against a database of roughly 500,000 entries, largely sourced from natural products and environmental toxicology literature. While this approach increases the apparent annotation rate, it carries significant risk, as the true accuracy of these assignments is unclear, and the authors offer no straightforward means of evaluating how many of these features are actually correct.
Metabolite identification remains one of the most challenging bottlenecks in modern metabolomics and is a step prone to critical errors that may go undetected. This problem is particularly acute in studies using liquid chromatography–mass spectrometry, where concerns about the validity of proposed metabolite identities are widespread. A recurring issue is the misidentification of metabolic features, sometimes to the point of implausibility, which are nevertheless put forward as potential biomarkers. Such poorly executed studies risk undermining confidence in metabolic phenotyping as a reliable approach for biomarker discovery, the characterization of metabolic perturbations, and the broader understanding of biological processes.
With respect to the specific annotations reported, the authors claim detection of the pesticides imidacloprid and propamocarb, as well as the microbially derived metabolites N-acetylmannosamine and phloroglucinol. Literature evidence indicates that both imidacloprid and propamocarb undergo extensive metabolism before excretion, making their direct detection in samples somewhat unexpected, though not entirely implausible. Supporting studies typically report these compounds either via targeted QQQ-MS assays or in individuals with direct occupational exposure, such as farm workers. To strengthen confidence in these annotations, it would be advisable to provide the corresponding MS/MS spectra and, ideally, confirm the assignments by spiking authentic standards of both pesticides prior to publication.
Regarding the detection of phloroglucinol, this metabolite is also typically metabolised extensively prior to excretion in urine, appearing mainly as the glucuronide and sulphated species. It is possible that the authors may have detected an in-source fragment from either of these species. It would be good to spike an authentic standard of phloroglucinol and match retention time to the feature of interest. Finally, what evidence is there that the authors detected N-acetylmannosamine, and not the more common isomers: N-acetylglucosamine or N-acetylgalactosamine? Again, a spiking experiment would be of interest. Given the potential for generating and propagating errors with this (and other such) softwares, I would advocate for stronger warning regarding assignment confidence and more emphasis in the manuscript around the limitations of this package.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have carefully read the authors' revised manuscript and their detailed point-by-point response letter. I appreciate the authors' thorough and constructive revisions addressing all of my previous comments.

In the revised version, the manuscript has been substantially improved in clarity, methodological transparency, and scientific rigor. All issues I raised in my initial review—including the accessibility of MetOriginDB, clarification of origin information resolution, explicit description of the underlying algorithms, differentiation from previous frameworks (e.g., MetDNA), and validation using public datasets—have been fully addressed.

Overall, I find that the authors have responded comprehensively to all concerns. I have no further comments. Thanks!

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I was asked to assess the authors' responses to Reviewer #2's Comments. The authors have appropriately addressed all the comments and concerns, performed additional experiments as requested in comments 3 and 4 of Reviewer #2 and modified the manuscript accordingly as highlighted. The manuscript overall is well written, critically assessed with reproducible methodology. Hence, I recommend its publication to Nature Communications.

(Remarks on code availability)
The installation instructions and documentation are easy to understand and well-maintained. I did not review the source code.

Reviewer Comments:
Reviewer #1 (Remarks to the Author):

I have reviewed the manuscript on TidyMass2. Compared with the first-generation TidyMass, the manuscript clearly describes the advanced functionalities of the updated framework, which makes the logic of the paper straightforward to evaluate. The authors highlight three main contributions: (1) the construction and utility of MetOriginDB, (2) an MS2-free module analysis pipeline, and (3) a user-friendly pipeline through Shiny apps. Among these, I focused my evaluation on points (1) and (2). Below, I provide comments and concerns regarding the manuscript.

Response: Thanks for your comprehensive review of TidyMass2 and for recognizing the methodological improvements beyond the original TidyMass framework. We have carefully addressed all your comments below.

Comment #1. Accessibility of MetOriginDB

The necessity of MetOriginDB is well understood, and this resource would be highly valuable to the broader community. However, it appears that the database is currently only provided as an R binary format (.rda). For third-party evaluation, it is essential that the authors also provide an ASCII-accessible format (e.g., Excel or CSV or TSV). Without such accessibility, independent validation of this resource is not feasible.

Response: We agree that providing MetOriginDB in multiple formats will enhance its utility for the broader research community and enable independent validation. We have now made the database available in CSV and TSV in addition to the original R binary format. All formats include complete metabolite information with source annotations and are available at https://www.tidymass.org/databases/. We also revised our manuscript to make it clearer for readers (line 772).

Comment #2. Resolution of origin information

For researchers studying the microbiome, origin information at the class or species level would be highly desirable. In this regard, MicrobeMASST currently appears to offer more granular classification. The authors should more clearly discuss the strengths and limitations of their approach in comparison. It is, of course, also true that metabolome information at the class or species level is not yet systematically available in public databases. Therefore, updating the discussion with the latest developments in microbiome metabolomics databases, explaining why such information cannot yet be incorporated, and describing the future directions of MetOriginDB would strengthen the manuscript.

Response: We appreciate your insightful comments regarding the granularity of origin information of metabolites in MetOriginDB. In response, we have substantially enhanced MetOriginDB to incorporate species-level annotations wherever available. The updated database now includes species-specific information for 56,949 microbiota-derived metabolites, with complete taxonomic hierarchies (phylum, class, order, family, genus, and species) using the same methods from MicrobeMASST[1]. This process is now explicitly documented in the **Methods** section of the revised manuscript (lines 776-779).

We also acknowledge that MicrobeMASST provides excellent species-level resolution through MS^2 spectral matching. This approach is limited to the subset of features with available MS^2 spectra[1]. Our framework complements this by enabling origin inference for all detected features with annotations by

any method. Furthermore, MetOriginDB encompasses seven distinct sources of metabolites (Human, animal, bacteria, drug, food, plant, and environment) beyond those only derived from the microbiota.

Future refinements of MetOriginDB could extend the level of granularity to metabolites from other sources. For example, for human-derived metabolites, their origin (such as blood, urine, or tissue) can be specified to provide a more precise source of the metabolites. In addition, all source information is derived from knowledge-based databases. In the future, this could be enhanced by incorporating additional origin information from public metabolomics repositories, applying a similar concept to that used by MicrobeMASST.

We have expanded the manuscript's discussion to clarify this complementary relationship and to outline future directions for improving taxonomic resolution (lines 498-503, 570-576).

Comment #3. Peak picking and alignment in TidyMass

It is not clear whether TidyMass (and TidyMass2) rely on the XCMS package for peak picking and alignment of raw MS data. This might have been described in the first-generation paper, but I recommend that the authors clarify this explicitly. If the algorithms were implemented from scratch, that should also be clearly stated.

Response: We apologize for the lack of clarity regarding the underlying algorithms used for data processing. TidyMass2 utilizes the well-established algorithm from the *xcms* R package for peak detection and alignment, specifically implementing the centWave algorithm, which has been extensively validated for LC-MS metabolomics data. We have now explicitly stated this in the **Methods** section in the revised manuscript (lines 920-924).

Comment #4. Similarity to MetDNA

The idea of the module-based functional analysis strongly resembles the previously reported MetDNA framework by the corresponding author, Xiaotao Shen. While there may be updated features in detection or processing, the conceptual foundation appears to be the same. It would be appropriate to cite the earlier study explicitly and emphasize what specific improvements TidyMass2 offers beyond MetDNA.

Response: We thank the reviewer for highlighting the conceptual link to our previous MetDNA work[2]. While both approaches leverage metabolic networks, TidyMass2 incorporates substantial methodological advances beyond MetDNA.

MetDNA is based on the principle that metabolites forming a reaction pair often share similar MS^2 spectra, enabling annotation through spectral similarity within the metabolic network. In contrast, TidyMass2 does not rely on this principle for metabolite annotation. Instead, it matches features primarily based on accurate mass (*m/z*), integrating retention time (RT), isotope patterns, and adduct information to cluster features into distinct metabolite classes. Redundancy within these classes is then reduced according to annotation confidence. Subsequently, TidyMass maps these curated annotations onto the metabolic network and applies graph-based community detection algorithms for module identification.

We have now explicitly acknowledged MetDNA and emphasized their differences in the **Methods** session of the revised manuscript (lines 815-822).

Comment #5. Differentiation from existing pipelines

While the utility of the TidyMass2 pipeline is clear, it remains somewhat unclear how its ideas and functionalities fundamentally differentiate it from existing pipelines. Experienced bioinformatics researchers might already be able to produce similar outputs without using TidyMass2. Explicitly articulating the unique aspects of TidyMass2 in terms of innovation and differentiation would make the contribution clearer.

Response: We appreciate the importance of clearly articulating how TidyMass2 advances beyond existing tools. To this end, we have provided a more detailed comparison between TidyMass2 and other platforms in the revised **Supplementary Table 2**. The unique contribution of the TidyMass2 framework lies in its integration of comprehensive metabolite origin inference with metabolic feature-based functional module analysis within a unified, reproducible workflow. Unlike platforms that focus on specific aspects of metabolomics analysis, TidyMass2 delivers an end-to-end solution that preserves data integrity through the "mass_dataset" object structure while offering both command-line flexibility and a user-friendly graphical interface.

We have expanded our comparative analysis in the revised **Supplementary Table 2** and emphasized these distinctions in the **Discussion** section of the revised manuscript to make our contribution clearer to readers (lines 540-544).

Comment #6. Validation on public datasets

If the authors wish to emphasize the utility of TidyMass2 (which I greatly respect), I recommend validating the pipeline not only on the in-house urine metabolomics dataset from human pregnancy but also on public datasets stored in repositories such as Metabolomics Workbench or MetaboLights. Since the current dataset originates from the authors' group, they are very familiar with its properties, which may bias the demonstration. In contrast, analyzing external datasets would allow the community to better assess whether the pipeline is truly "globally applicable." For example: (A) Can the pipeline robustly handle diverse raw MS datasets without errors, as MS-DIAL and MZmine have demonstrated? (B) Through re-analysis of public datasets, to what extent can TidyMass2 extract new biological insights that were not reported in previous studies?

Response: We appreciate this constructive suggestion to demonstrate TidyMass2's broader applicability. According to your suggestions, we have performed a comprehensive re-analysis of two additional public datasets from MetaboLights (ID: MTBLS5828)[3] and Metabolomics Workbench (ID: PR000918)[4] from different biological contexts.

Case study 2: Chronic Inflammatory Demyelinating Polyradiculoneuropathy (CIDP) metabolomics study.

This case study is based on a previously published metabolomics analysis of metabolic changes in Chronic Inflammatory Demyelinating Polyradiculoneuropathy (CIDP) patients[3]. We retrieved the raw metabolomics data from MetaboLights (ID: MTBLS5828) and processed it using TidyMass2 for peak picking and data cleaning. The original study used conventional approaches to identify and quantify metabolites, perform differential analyses between the CIDP and control groups, and then conduct pathway enrichment separately for positive and negative ion modes. In total, they reported 14 enriched pathways in positive mode and 20 in negative mode using the KEGG database. Because no multiple-testing correction was applied in the original enrichment analysis, we recalculated adjusted *p-values* (n = 200, corresponding to the approximate number of KEGG pathways). Eight pathways reached statistical significance (FDR < 0.05): Bile secretion, Cholesterol metabolism, Purine metabolism, Aldosterone-regulated sodium reabsorption, Steroid hormone biosynthesis, Primary bile acid

biosynthesis, Metabolic pathways, and Caffeine metabolism. As "Metabolic pathways" is a very broad category in KEGG, we excluded it, yielding seven final pathways consistent with the original study.

We then applied TidyMass2's metabolic feature-based functional module analysis (FFMA) to the differential features (1,847 total), identifying 15 functional modules with FDR < 0.05 (**Supplementary Figure 8a**). As expected, most pathways reported by the traditional method mapped to FFMA modules. For example, Steroid hormone biosynthesis is mapped to Module 2 (steroid hormone metabolism), Purine metabolism to Module 4, and Bile secretion to Module 14 (bile acid metabolism) (**Supplementary Figure 8b-d**).

Notably, FFMA also revealed modules not detected by the traditional approach. For instance, Module 3 corresponded to a core carbohydrate and energy metabolism module encompassing pathways such as Fructose and mannose metabolism, Galactose metabolism, Glucagon signaling, Starch and sucrose metabolism, AMPK signaling, Glycolysis/Gluconeogenesis, and Phenylalanine-Tyrosine-Tryptophan biosynthesis (**Supplementary Figure 8e**). This module integrates central processes of energy production, redox balance, biosynthetic flux, and energy sensing. Activated immune cells (macrophages, T cells) engaged in demyelination undergo metabolic reprogramming toward glycolysis, the pentose phosphate pathway, and anabolic metabolism to fuel their inflammatory and phagocytic functions[5,6], while Schwann cells rely on intact AMPK signaling for proper myelination and energy homeostasis[7]. Moreover, altered purinergic signaling (ATP/adenosine) and tryptophan-kynurenine metabolism are recognized modulators of neuroimmune cross-talk, and impaired lysosomal clearance of myelin debris can perpetuate inflammation and impede remyelination.

Case study 3: Longitudinal pregnancy plasma metabolomics study.

This case study utilizes previously published metabolomics data from a human pregnancy longitudinal study[4]. Raw metabolomics data (Metabolomics Workbench ID: PR000918) were processed using the TidyMass2 framework, encompassing all steps from peak picking to biological interpretation. This analysis identified 1,330 significantly dysregulated metabolic features, with 683 showing an increase and 647 a decrease during pregnancy. To further investigate, metabolic feature-based functional module analysis (FFMA) was applied to these changing features, revealing 17 dysregulated metabolic modules with a *p-value* < 0.05 (**Supplementary Figure 9a**).

Importantly, most of the pathways reported in the original study were also recovered in our analysis. For example, Aminoacyl-tRNA biosynthesis, Tyrosine metabolism, and Phenylalanine metabolism are mapped to Module 1 (amino acid metabolism), Fatty acid biosynthesis, Linoleic acid metabolism, and Arachidonic acid metabolism to Module 4 (lipid metabolism), Amino sugar and nucleotide sugar metabolism to Module 10, and Steroid hormone biosynthesis to Module 11 (**Supplementary Figure 9b-e**).

Modules that were not detected by traditional methods were also revealed by FFMA in this case study. For example, Module 2 is associated with carbohydrate transport and utilization and includes pathways such as Galactose metabolism, Fructose and mannose metabolism, ABC transporters, and Carbohydrate digestion and absorption (**Supplementary Figure 9f**). This module combines dietary sugar processing, intestinal absorption, and systemic transport of monosaccharides, making a positive impact on maternal-fetal energy supply. During pregnancy, maternal glucose homeostasis undergoes marked adaptations, with increased carbohydrate fluxes to support rapid fetal growth and placental function[8]. In particular, Galactose and mannose metabolism are crucial for glycosylation reactions underlying hormone regulation and fetal tissue development[9], while ABC transporters mediate nutrient exchange across

placental barriers[10]. And previous research has demonstrated that dysregulation of maternal carbohydrate handling is associated with gestational insulin resistance and gestational diabetes mellitus, conditions that can impair fetal growth trajectories and long-term offspring metabolic health[11,12].

Similarly, Module 12 is another example that was detected by FFMA and represents a module centered on the metabolism of xenobiotics by cytochrome P450 (**Supplementary Figure 9g**). Cytochrome P450 enzymes mediate the clearance of drugs, dietary compounds, and environmental chemicals[13]. During pregnancy, the hormonal and relevant physiological changes are known to alter CYP450 activity[14,15], thereby influencing maternal drug metabolism, modifying the clearance of xenobiotics, and potentially affecting the extent of fetal exposure. In summary, these examples highlight TidyMass2's ability to uncover novel biological insights beyond existing findings.

These two additional case studies further showcase TidyMass2's capability to robustly process diverse raw MS datasets and reveal new biological insights. All findings have been incorporated into our updated Supplementary Information (**Supplementary Note** and **Supplementary Figures** 8a-e, 9a-g) and revised manuscript (lines 475-478).

References

1. Zuffa, S. et al. microbeMASST: a taxonomically informed mass spectrometry search tool for microbial metabolomics data. Nat. Microbiol. 9, 336–345 (2024).
2. Shen, X. et al. Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. Nat. Commun. 10, 1516 (2019).
3. Fu, J. et al. Metabolic disorder and intestinal microflora dysbiosis in chronic inflammatory demyelinating polyradiculoneuropathy. Cell Biosci. 13, 6 (2023).
4. Liang, L. et al. Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. Cell 181, 1680-1692.e15 (2020).
5. Pålsson-McDermott, E. M. & O'Neill, L. A. J. Targeting immunometabolism as an anti-inflammatory strategy. Cell Res. 30, 300–314 (2020).
6. Soto-Heredero, G., Gómez de las Heras, M. M., Gabandé-Rodríguez, E., Oller, J. & Mittelbrunn, M. Glycolysis – a key player in the inflammatory response. FEBS J. 287, 3350–3369 (2020).
7. Liu, X. et al. AMPK Negatively Regulates Peripheral Myelination via Activation of c-Jun. Mol. Neurobiol. 54, 3554–3564 (2017).
8. Angueira, A. R. et al. New Insights Into Gestational Glucose Metabolism: Lessons Learned From 21st Century Approaches. Diabetes 64, 327–334 (2015).
9. He, M., Zhou, X. & Wang, X. Glycosylation: mechanisms, biological functions and clinical implications. Signal Transduct. Target. Ther. 9, 194 (2024).
10. Joshi, A. A. et al. Placental ABC Transporters: Biological Impact and Pharmaceutical Significance. Pharm. Res. 33, 2847–2878 (2016).
11. Murray, S. R. & Reynolds, R. M. Short- and long-term outcomes of gestational diabetes and its treatment on fetal development. Prenat. Diagn. 40, 1085–1091 (2020).
12. Parrettini, S., Caroli, A. & Torlone, E. Nutrition and Metabolic Adaptations in Physiological and Complicated Pregnancy: Focus on Obesity and Gestational Diabetes. Front. Endocrinol. 11, 611929 (2020).
13. Hakkola, J., Hukkanen, J., Turpeinen, M. & Pelkonen, O. Inhibition and induction of CYP enzymes in humans: an update. Arch. Toxicol. 94, 3671–3722 (2020).
14. Isoherranen, N. & Thummel, K. E. Drug Metabolism and Transport During Pregnancy: How Does Drug Disposition Change during Pregnancy and What Are the Mechanisms that Cause Such Changes? Drug Metab. Dispos. 41, 256–262 (2013).

15. Fashe, M. M. et al. Impact of pregnancy related hormones on drug metabolizing enzyme and transport protein concentrations in human hepatocytes. Front. Pharmacol. 13, 1004010 (2022).

Reviewer #2 (Remarks to the Author):

The article: "TidyMass2: Advancing LC-MS Untargeted Metabolomics Through Metabolite Origin Inference and Metabolic Feature-based Functional Module Analysis" by Wang and coworkers, presents an original software package for metabolic feature annotation and determination of analyte origin.

Response: Thanks for your comprehensive review of our manuscript and thoughtful suggestions. We have carefully addressed all your comments below.

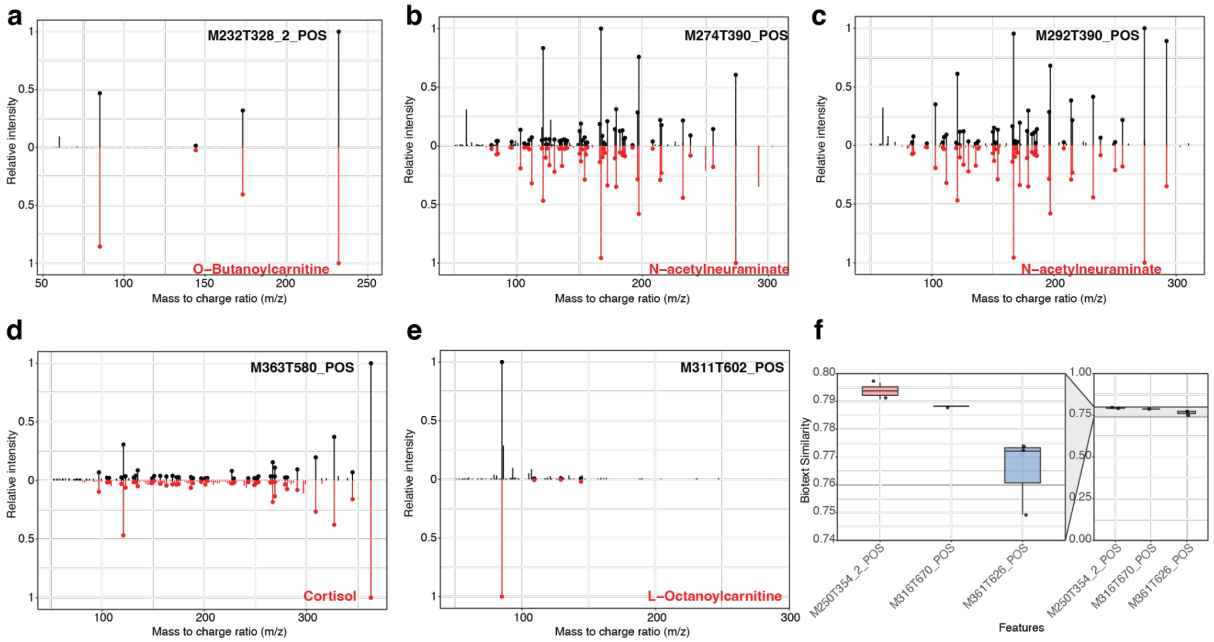
Comment #1. The authors describe a rise in "interpretable" features from 5.8% to 58.8%, achieved by abandoning MS^2-based annotations from matched fragmentation spectra in favor of MS^1 accurate m/z matching against a database of roughly 500,000 entries, largely sourced from natural products and environmental toxicology literature. While this approach increases the apparent annotation rate, it carries significant risk, as the true accuracy of these assignments is unclear, and the authors offer no straightforward means of evaluating how many of these features are actually correct.

Response: We thank the reviewer for raising this point about annotation accuracy in our metabolic feature-based functional module analysis (FFMA).

First, because our case study involved human data, the metabolic network we used was derived from humans and contains 9,630 metabolites with valid HMDB and KEGG identifiers (**Fig. 3b**). Accordingly, **we did not match features against the entire MetOriginDB (~500,000 compounds), but instead against this curated human sub-database (9,630 metabolites).** We are sorry for the confusion, and have revised the manuscript to clarify this for readers (lines 281-283, 836-838).

Second, in the case study, we used our in-house library (provided by the case study paper's authors) to validate features annotated by FFMA. In total, eight features were identified. Among these, five were correctly annotated by FFMA (Two features are the same metabolites) (**Supplementary Figure 6a-e**). For the remaining three features, we compared the biological similarity between the true annotations and the FFMA annotations and found high biological similarity (**Supplementary Figure 6f**). These results demonstrate that, although three features were not perfectly annotated, the FFMA still assigned biologically consistent annotations. Because FFMA is designed to extract functional modules, even features with slightly inaccurate individual annotations but similar functional profiles still contribute to the same or related modules.

We have included these results in the revised manuscript (lines 413-416, 893-902) and **Supplementary Information** and have highlighted this point in the **Discussion** section of the revised manuscript (lines 517-520).



Supplementary Figure 6 | Validation of FFMA-annotated features using in-house library and biological similarity analysis. a-e, Comparison of MS^2 spectra with the library for the 5 matched features: **a,** M232T328_2_POS; **b,** M274T390_POS; **c,** M292T390_POS; **d,** M363T580_POS; **e,** M311T602_POS. **f,** Biological similarity between the FFMA annotations and the true annotations for the 3 unmatched features.

Comment #2. Metabolite identification remains one of the most challenging bottlenecks in modern metabolomics and is a step prone to critical errors that may go undetected. This problem is particularly acute in studies using liquid chromatography–mass spectrometry, where concerns about the validity of proposed metabolite identities are widespread. A recurring issue is the misidentification of metabolic features, sometimes to the point of implausibility, which are nevertheless put forward as potential biomarkers. Such poorly executed studies risk undermining confidence in metabolic phenotyping as a reliable approach for biomarker discovery, the characterization of metabolic perturbations, and the broader understanding of biological processes.

Response: We fully agree with the reviewer that metabolite identification remains one of the most challenging steps in LC-MS metabolomics studies. In TidyMass[1] and TidyMass2, we did not introduce a new annotation algorithm; instead, we implemented the most widely used and transparent approaches for metabolite annotation in this field, namely based on (1) accurate mass (*m/z*), (2) MS^2 spectra, and (3) retention time (RT), and the annotations are clearly labeled according to the Metabolomics Standards Initiative (MSI) guidelines[2]. Specifically, level 1 annotations are based on in-house reference libraries analyzed on the same LC system (*m/z* + RT + MS^2); level 2 uses MS^2 libraries (*m/z* + MS^2); and level 3 relies on accurate mass (*m/z*) only. These annotation levels are explicitly provided in the TidyMass2 annotation output and described in detail in our original TidyMass publication[1]. For the present manuscript, we have used exactly these standard workflows and have not added new annotation methods.

To address the reviewer's comment and avoid any misunderstanding, we have emphasized this annotation strategy more explicitly in the **Methods** section of the revised manuscript (lines 932-939).

Comment #3. With respect to the specific annotations reported, the authors claim detection of the pesticides imidacloprid and propamocarb, as well as the microbially derived metabolites N-acetylmannosamine and phloroglucinol. Literature evidence indicates that both imidacloprid and propamocarb undergo extensive metabolism before excretion, making their direct detection in samples

somewhat unexpected, though not entirely implausible. Supporting studies typically report these compounds either via targeted QQQ-MS assays or in individuals with direct occupational exposure, such as farm workers. To strengthen confidence in these annotations, it would be advisable to provide the corresponding MS/MS spectra and, ideally, confirm the assignments by spiking authentic standards of both pesticides prior to publication.

Response: We thank you for raising this point about the annotation of imidacloprid and propamocarb in our case study.

Both compounds, as noted, undergo extensive biotransformation in the human body. Nevertheless, several studies have demonstrated their presence in human urine from the general population. For instance, Tao *et al.* detected imidacloprid in human urine using a triple quadrupole mass spectrometry (QQQ-MS) assay[3], and Wang *et al.* reported imidacloprid in 100% of rural and 95% of urban urine samples (n = 295) in China using the same analytical approach[4]. Similarly, Ottenbros *et al.* conducted a large-scale suspect-screening study of adult and child urine and detected propamocarb in urine samples[5], with an inverse association between urinary propamocarb levels and organic produce intake, suggesting dietary exposure as a potential source.

Following the reviewer's recommendation, we performed validation experiments using authentic standards of both imidacloprid and propamocarb. **Because this case study originates from a published dataset from Stanford University**[6], it was not possible for us to reproduce the exact LC conditions (column, mobile phase, and other chromatographic parameters). Therefore, we validated our findings by comparing the *m/z* and MS^2 spectra of experimental features with those of the authentic standards (**Supplementary Figure 4a**), rather than matching retention times. This approach corresponds to a Level 2 annotation according to the Metabolomics Standards Initiative (MSI)[2], which we have now clarified in the revised manuscript to make it clear for readers.

All corresponding results have been incorporated into the revised manuscript (lines 349-351) and the **Supplementary Information**.

Comment #4. Regarding the detection of phloroglucinol, this metabolite is also typically metabolised extensively prior to excretion in urine, appearing mainly as the glucuronide and sulphated species. It is possible that the authors may have detected an in-source fragment from either of these species. It would be good to spike an authentic standard of phloroglucinol and match the retention time to the feature of interest. Finally, what evidence is there that the authors detected N-acetylmannosamine, and not the more common isomers: N-acetylglucosamine or N-acetylgalactosamine? Again, a spiking experiment would be of interest.

Response: We thank you for raising this point regarding the annotation of phloroglucinol and N-acetylmannosamine in the case study.

With respect to phloroglucinol metabolism, as noted, it undergoes biotransformation in the human body. And as you mentioned, several studies have reported its detection in plasma using QQQ-MS assays[7]. We didn't find publications that have reported the detection of phloroglucinol in human urine samples using LC-MS directly. However, other analytical approaches have been described. For example, Nascimento dos Santos *et al.* developed a photoelectrochemical method to detect phloroglucinol in human urine samples[8].

Following the reviewer's recommendation, we conducted additional validation experiments using authentic standards of both phloroglucinol and N-acetylmannosamine. As noted in our response to

Comment #4, because this case study originates from a published dataset from Stanford University[6], it was not possible to exactly reproduce the original LC conditions (column type, mobile phase, and other chromatographic parameters). Instead, we validated our findings by comparing the *m/z* and MS^2 spectra of experimental features with those of authentic standards (**Supplementary Figure 4a**). This approach corresponds to a Level 2 annotation according to the MSI^2, which we have now clarified in the revised manuscript to make it clear for readers (lines 359-361).

For N-acetylmannosamine, we agree that without ion mobility separation or other additional orthogonal methods, it is difficult to unequivocally distinguish it from N-acetylglucosamine or N-acetylgalactosamine. This is also why our annotation represents a Level 2 identification according to MSI^2.

We have acknowledged this as a limitation in the **Discussion** section of the revised manuscript (lines 547-550).

All of these new results have been added to the revised manuscript (lines 359-361) and the **Supplementary Information**.

Comment #5. Given the potential for generating and propagating errors with this (and other such) softwares, I would advocate for stronger warning regarding assignment confidence and more emphasis in the manuscript around the limitations of this package.

Response: We fully agree on the critical importance of transparently communicating the confidence levels and limitations of our metabolic feature-based functional module analysis. Our method, just like other computation methods, such as MetDNA[9], Mummichog[10], and PIUMet[11], should be the level 3 annotations according to MSI^2. We have now incorporated annotation confidence indicators (Level 3)[2] into all annotation outputs of the analysis (metabolic feature-based functional module analysis) and have updated terminology throughout to more accurately reflect uncertainty.

In addition, we have substantially revised the manuscript to acknowledge the limitations of our method and to clearly state that network-based annotations are putative and require experimental validation (lines 579-582).

References

1. Shen, X. et al. TidyMass an object-oriented reproducible analysis framework for LC–MS data. Nat. Commun. 13, 4365 (2022).
2. Sansone, S.-A. et al. The Metabolomics Standards Initiative. Nat. Biotechnol. 25, 846–848 (2007).
3. Tao, Y. et al. Urinary monitoring of neonicotinoid imidacloprid exposure to pesticide applicators. Sci. Total Environ. 669, 721–728 (2019).
4. Wang, L. et al. Occurrence and Profile Characteristics of the Pesticide Imidacloprid, Preservative Parabens, and Their Metabolites in Human Urine from Rural and Urban China. Environ. Sci. Technol. 49, 14633–14640 (2015).
5. Ottenbros, I. B. et al. Urinary pesticide mixture patterns and exposure determinants in the adult population from the Netherlands and Switzerland: Application of a suspect screening approach. Environ. Res. 239, 117216 (2023).
6. Shen, X. et al. Longitudinal urine metabolic profiling and gestational age prediction in human pregnancy. Brief. Bioinform. 26, bbaf059 (2025).
7. Kim, H. et al. Determination of phloroglucinol in human plasma by high-performance liquid chromatography–mass spectrometry. J. Chromatogr. B 792, 307–312 (2003).

8. dos Santos, J. R. N. et al. Photoelectrochemical determination of phloroglucinol based on a combination of a ceramic perovskite and bismuth vanadate. Microchim. Acta 191, 609 (2024).
9. Shen, X. et al. Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. Nat. Commun. 10, 1516 (2019).
10. Li, S. et al. Predicting Network Activity from High Throughput Metabolomics. PLOS Comput. Biol. 9, e1003123 (2013).
11. Pirhaji, L. et al. Revealing disease-associated pathways by network integration of untargeted metabolomics. Nat. Methods 13, 770–776 (2016).