

MeTime: an R package for reproducible longitudinal metabolomics data analysis

Bharadwaj Marella^{1,✉}, Patrick Weinisch^{1,✉}, Lara Vehovec¹, Vinh Tran¹, Josef J. Bless¹,
Yacoub A. Njipouombe Nsangou¹, Gabi Kastenmüller¹, Matthias Arnold^{1,2,✉} 

¹Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

²Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

✉Equal contribution.

*Corresponding author. Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA. E-mail: matthias.arnold@helmholtz-munich.de
Associate Editor: Reema Singh

Abstract

Summary: MeTime is an opensource R package for reproducible analysis of longitudinal metabolomics data. It builds upon a central S4 container, `metime_analyser`, that stores multiple datasets, associated metadata and analysis outputs, enabling unified handling of complex longitudinal studies. Analyses are constructed by piping modular functions, beginning with data transformations (`mod_*`), followed by calculations (`calc_*`), and optional meta-analysis (`meta_*`), so entire workflows remain transparent and easy to modify. MeTime wraps numerous existing methods within a consistent interface, including sample and metabolite distributions, correlation/distance matrices, dimensionality reduction (PCA, UMAP, t-SNE), random forest imputation and feature selection via Boruta, eigenmetabolites and WGCNA-based clustering, conservation index analysis, regression models (linear, mixed-effects, and generalized additive), and partial-correlation networks. By retaining all intermediate results and provenance within the container, MeTime facilitates iterative exploration and ensures reproducible reporting via automatically generated HTML/PDF outputs. Comprehensive user guides, case studies and reference documentation accompany the package, making MeTime a versatile platform for longitudinal omics workflows.

Availability and implementation: MeTime is available as an open-source R package and can be installed directly from GitHub (<https://github.com/compneurobio/MeTime>). Source code, installation instructions, documentation, tutorials, and reproducible case-study workflows are provided within the repository. MeTime has been tested on Microsoft Windows, Unix/Linux, and macOS. For users requiring a containerized environment, a Docker implementation is additionally provided through the GitHub Container Registry.

1 Introduction

Longitudinal metabolomics provides a powerful framework for characterizing temporal changes in metabolite profiles within an organism. By capturing within-subject dynamics over time, longitudinal designs enable the investigation of molecular responses to stimuli, interventions, and disease progression, while accounting for inter-individual variability. Metabolomic profiles provide a downstream readout of biochemical variation shaped by molecular, environmental, lifestyle, and disease-related influences. Longitudinal measurements can capture dynamic changes associated with disease progression, intervention response, and predict future health outcomes, including all-cause mortality (Lacruz *et al.* 2018, Rosato *et al.* 2018, Arnold *et al.* 2020, Castelli *et al.* 2022, Shen *et al.* 2023).

Advances in mass spectrometry technologies and data acquisition workflows have led to a rapid increase in the generation of longitudinal metabolomics datasets. This growth has intensified the demand for scalable, flexible, and reproducible computational approaches capable of handling high-dimensional, multi-timepoint data (Wishart *et al.* 2022, Trifonova *et al.* 2023). The analysis of such data commonly relies on a combination of multivariate methods, regression-based approaches (Gibbons *et al.* 2010), conservation index analysis (Yousri *et al.* 2014), and network-based models to capture temporal structure and inter-metabolite relationships (Krumstiek *et al.* 2011, Epskamp *et al.* 2018).

In parallel, the metabolomics community has placed increasing emphasis on reproducibility, transparency, and open science practices (Considine *et al.* 2017, Du *et al.* 2022). Contemporary

Received: 12 May 2026. Revised: 1 August 2026. Accepted: 26 August 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

publication standards often require the sharing of source code, parameters, and intermediate results alongside manuscripts (Baker 2016). While these practices improve scientific rigor, they also impose a substantial burden on researchers, who must design analysis pipelines that are both flexible and fully traceable.

Several toolboxes support metabolomics data preprocessing, statistical analysis, visualization or pathway interpretation, such as MetaboAnalyst, maplet, and MeTEor (Chong *et al.* 2018, Chetnik *et al.* 2022, Grabert *et al.* 2024); however, these tools were generally not designed around the specific requirements of longitudinal metabolomics studies involving repeated measurements, multiple datasets, heterogeneous platforms and the need to preserve analysis provenance across sequential workflows (Supplementary Table 1, available at [supplementary material Bioinformatics Advances](#) online). Here, we present MeTime, an open-source R package (<https://github.com/compneurobio/MeTime/>) that brings together longitudinal metabolomics analyses in a single unified framework: it stores multiple datasets in a central S4 object, wraps a range of existing analysis methods behind a consistent interface, supports modular pipeline construction, and generates self-contained reports for transparent and reproducible workflows and results. In addition to describing the software design and functionality, we demonstrate the use of MeTime through an executable longitudinal metabolomics case study, illustrating how complex analytical pipelines and their results can be generated, stored, and reported in a fully reproducible manner.

2 Implementation and functionality

MeTime enables transparent and reproducible analytical pipelines for longitudinal omics data by centering all inputs, outputs, and provenance in a single S4 container. This object stores raw and processed metabolite abundance data from both targeted and untargeted metabolomics platforms, analytical results, and detailed metadata protocoling each pipeline step (called functions, parameter settings, and execution order). Pipelines are assembled by sequentially passing the container through S4 methods using the native R pipe operator (`|>`) or `%>%` from magrittr (Bache *et al.* 2022), making analyses straightforward to review, reproduce, and retrace.

A key design goal is native support for longitudinal metabolomics projects that include multiple datasets (e.g. cohorts, experimental batches, or analytical platforms) without requiring full sample overlap. While existing container classes such as SummarizedExperiment (Morgan *et al.* 2023) and data structures implemented in maplet (Chetnik *et al.* 2022) provide robust solutions for single datasets, they do not directly address coordinated workflows spanning multiple heterogeneous datasets. MeTime addresses this by enabling independent yet synchronized analyses across datasets while storing all derived outputs within the data container (Fig. 1). Meta-analysis summaries are stored separately in a second dedicated class object, which cleanly separates primary results within one container or across two containers from downstream meta-analysis. See [Supplementary Information](#),

available at [supplementary material Bioinformatics Advances](#) online for more details on the architecture of these classes.

MeTime follows a modular philosophy in which each analytical component (e.g. feature selection, regressions, data driven networks) is implemented as an independent function that can be combined into complete workflows. Functions follow a standardized tagging scheme reflecting their role in the pipeline (Fig. 1), enabling users to insert, remove, or reorder steps with minimal effort and support rapid prototyping. The package currently comprises ~80 functions spanning data loading, extraction, modification, augmentation, statistical analysis, meta-analysis of summary statistics, and export. Statistical methods include missing value imputation, dimensionality reduction, linear models, linear mixed-effects models, generalized additive mixed models, conservation index analysis, and data driven networks, alongside utilities for publication-ready visualizations and structured result tables. Although developed for metabolomics, the underlying architecture is extensible to other omics modalities. Templates and extensive documentation are provided to facilitate extension with custom functions while preserving compatibility with the core architecture (<https://github.com/compneurobio/MeTime/>).

Following pipeline execution, results remain embedded in the data container with the option to generate comprehensive PDF or HTML reports via knitr and Rmarkdown packages (Allaire *et al.* 2023). Reports enumerate pipeline steps in the order executed, including function names, parameter settings, and associated tables/figures (Fig. 1), and can be exported for sharing and archival. By consolidating results with code provenance and parameterization in a single document, MeTime supports transparent reporting and reproducibility.

3 Exemplary analysis

To demonstrate the functionality of MeTime, we provide two complementary usage examples. First, the package includes the publicly available HuMet dataset (Weinisch *et al.* 2024), which is used in the GitHub tutorials and example workflows as a small and accessible dataset for learning the package and exploring its main functions along with a complete walkthrough of the different analyses including dimensionality reduction, feature selection, conservation index analysis, association analysis, network analysis and cluster identification that can be conducted using MeTime in a dedicated Rmarkdown file ([Supplementary Information](#), available at [supplementary material Bioinformatics Advances](#) online). Second, we provide a fully reproducible workflow on GitHub (<https://github.com/compneurobio/MeTime/tree/main/docs/case-studies>) that replicates the association analyses using a controlled-access longitudinal metabolomics dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) measured with the targeted Biocrates MXP® Quant 500 platform from Marella *et al.* 2025. After quality control and construction of the data container, this analysis can be rerun directly using the provided code. The analyses resulting from [Supplementary Information](#) are provided in the [Supplementary Data](#), available at [supplementary material Bioinformatics Advances](#) online as report files, so that each code snippet can be viewed both as code and as its resulting output.

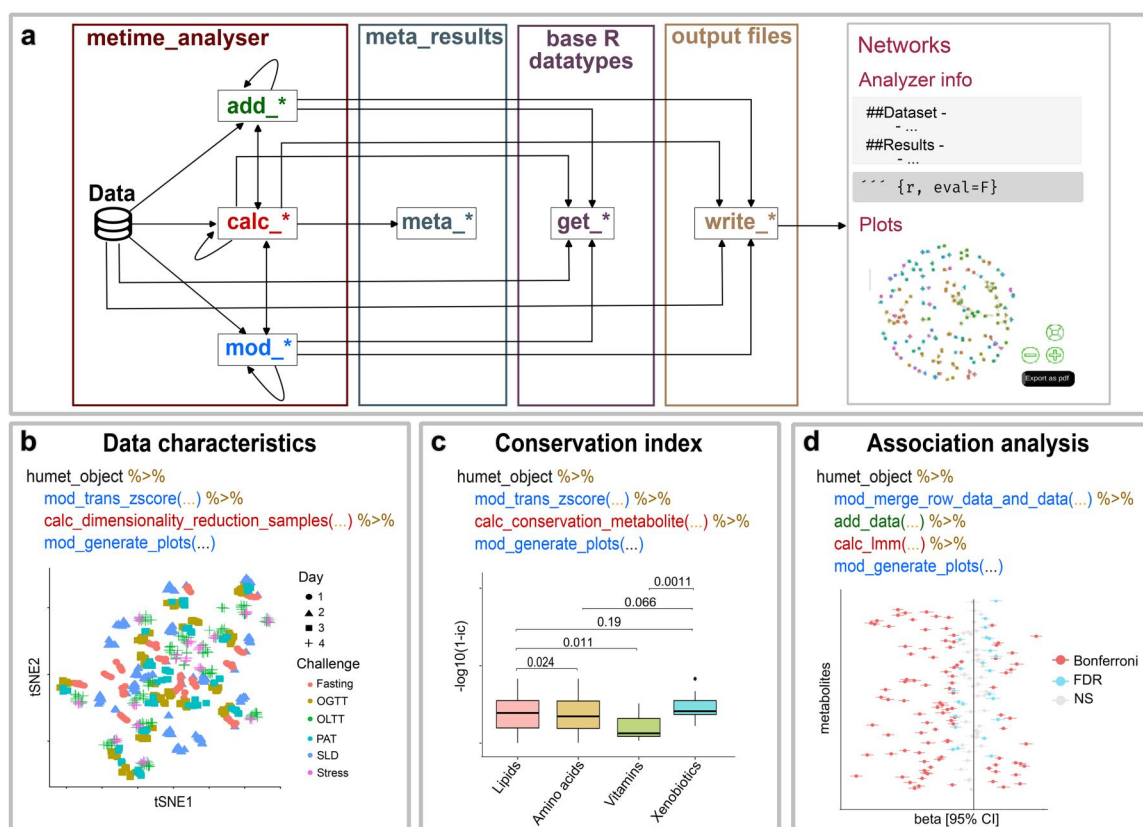


Figure 1 Overview of MeTime's architecture and example analyses. (a) A schematic of constructing a pipeline using MeTime: the central `metime_analyser` object holds input data and results; functions prefixed as `add_*` import or merge data, `mod_*` modify and annotate, `calc_*` perform analyses, `meta_*` compare summary statistics, `get_*` retrieves contents, and `write_*` generates reports. Arrows indicate, which function family can be applied consecutively. Summary statistics across two calculations or `metime_analyser` objects can be summarised in secondary `meta_results` objects and visualised via built-in plotting mechanisms. The remaining panels showcase examples on how the user can build such pipelines. (b) Data characteristics: an example pipeline that z-scores metabolite abundances, computes a *t*-SNE embedding, and plots samples by challenge and day of challenge. (c) Conservation index: box plots showing stability of metabolic profiles across lipid and non-lipid classes after transformation and conservation index calculation. (d) Association analysis: effect-size estimates and 95% confidence intervals for selected metabolites from a linear mixed-effects model; colors indicate significance after Bonferroni, FDR adjustment ($q < 0.05$) and nominal ($P < 0.05$) thresholds.

Together, these examples show that MeTime can be used both as an easy starting point for new users and as a reproducible framework for a wide range of longitudinal omics analyses.

4 Limitations

MeTime is not intended to replace available tools such as `maplet` but rather to complement and add missing components for longitudinal data analysis. Therefore, current limitations include limited preprocessing functionality beyond export/import functions to interface with these tools. Furthermore, some analyses as implemented in MeTime are computationally intensive. Although we provide support for a popular high-performance computing environment (SLURM), integration with workflow managers such as `targets` is currently not included but will be provided in future releases. Finally, while the most commonly used longitudinal metabolomics analysis methods are available in MeTime, it does not cover all possible statistical methods. However, the framework is modular and extensible, with

developer documentation provided for users who wish to add methods for their own applications.

5 Conclusion

MeTime provides a comprehensive framework for the reproducible analysis of longitudinal metabolomics data by enabling rapid construction of transparent and modular analytical pipelines. By integrating data management, statistical modelling, results storage, and report generation within an S4-based architecture, MeTime thereby reduces the technical overhead associated with longitudinal data analysis, supporting both methodological development and transparent scientific reporting.

Author contributions

Conceptualization & Methodology: GK, MA; Code development & software: BM, PW, MA; Funding acquisition: GK, MA; Testing: BM,

PW, VT, LV, YN, MA; Writing—original draft: BM, JJB, MA; Writing—review & editing: All authors.

Supplementary material

Supplementary material is available at *Bioinformatics Advances* online.

Conflicts of interest

None declared.

Funding

This work was supported by the National Institutes of Health/ the National Institute on Aging through grants 1R1AG057452, R01AG069901, U01AG061359, and R01AG081322. This work was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—HyperMet [FOR 5795; Grant number: 536691227] and by the German Federal Ministry of Education and Research (BMBF) [BiomarkKid, 01EA2203B and PlantIntake, 01EA2204B] under the umbrella of the European Joint Programming Initiative “A Healthy Diet for a Healthy Life” (JPI HDHL) and the ERA-NET Cofund ERA-HDHL [GA N° 696295] of the EU Horizon 2020 Research and Innovation Programme.

Data availability

No new primary data were generated in this study. The HuMet dataset used for tutorials and example analyses is publicly available through the HuMet Repository (<https://www.humet.org>) and is distributed with MeTime for demonstration purposes. Additional sample information and spectral data are available through MetaboLights under accession MTBLS89. The Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal metabolomics data used in supplementary case studies is available through the AD Knowledge Portal (doi:10.7303/9618123), associated clinical and demographic data is available through the ADNI data-sharing platform at LONI. The MeTime source code, documentation, case study presented here, and reproducible analysis workflows are openly available at <https://github.com/compneurobio/MeTime/>. Scripts and corresponding reports underlying the examples presented in this study are provided as [Supplementary Data](#).

References

- Allaire J, Xie Y, Dervieux C *et al.* 2023. rmarkdown: Dynamic documents for R. R package version 2.25. <https://github.com/rstudio/rmarkdown>.
- Arnold M, Nho K, Kueider-Paisley A *et al.* Sex and APOE ε4 genotype modify the Alzheimer's disease serum metabolome. *Nat Commun* 2020;**11**:1148.
- Bache SM, Wickham H, Henry L *et al.* 2022. Package ‘magrittr’. R. Package Version.
- Baker M. Why scientists must share their research code. *Nature* 2016. <https://doi.org/10.1038/nature.2016.20504>
- Castelli FA, Rosati G, Moguet C *et al.* Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. *Anal Bioanal Chem* 2022;**414**:759–89.
- Chetnik K, Benedetti E, Gomari DP *et al.* maplet: an extensible R toolbox for modular and reproducible metabolomics pipelines. *Bioinformatics* 2022;**38**:1168–70.
- Chong J, Soufan O, Li C *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 2018;**46**:W486–94.
- Considine EC, Thomas G, Boulesteix AL *et al.* Critical review of reporting of the data analysis step in metabolomics. *Metabolomics* 2017;**14**:7–16.
- Du X, Aristizabal-Henao JJ, Garrett TJ *et al.* A checklist for reproducible computational analysis in clinical metabolomics research. *Metabolites* 2022;**12**:87.
- Epskamp S, Waldorp LJ, Möttus R *et al.* The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behav Res* 2018;**53**:453–80.
- Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. *Annu Rev Clin Psychol* 2010;**6**:79–107.
- Grabert G, Dehncke D, More T *et al.* MeTEor: an R shiny app for exploring longitudinal metabolomics data. *Bioinform Adv* 2024;**4**:vbae178.
- Krumsiek J, Suhre K, Illig T *et al.* Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol* 2011;**5**:21–16.
- Lacruz ME, Kluttig A, Tiller D *et al.* Instability of personal human metabolome is linked to all-cause mortality. *Sci Rep* 2018;**8**:9810.
- Marella B, Weinisch P, Bless JJ *et al.* A seven-year longitudinal study of the Alzheimer's disease blood metabolome. *medRxiv* 2025.
- Morgan M, Obenchain V, Hester J *et al.* SummarizedExperiment: summarizedexperiment container. <https://doi.org/10.18129/B9.bioc.SummarizedExperiment>. 2023.
- Rosato A, Tenori L, Cascante M *et al.* From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics* 2018;**14**:37–20.
- Shen S, Zhan C, Yang C *et al.* Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. *Mol Plant* 2023;**16**:43–63.
- Trifonova OP, Maslov DL, Balashova EE *et al.* Current state and future perspectives on personalized metabolomics. *Metabolites* 2023;**13**:67.
- Weinisch P, Raffler J, Römisch-Margl W *et al.* The HuMet Repository: watching human metabolism at work. *Cell Rep* 2024;**43**:114416.
- Wishart DS, Guo A, Oler E *et al.* HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res* 2022;**50**:D622–31.
- Yousri NA, Kastenmüller G, Gieger C *et al.* Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics* 2014;**10**:1005–17.

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Bioinformatics Advances, 2026, 6, 1–4

<https://doi.org/10.1093/bioadv/vbag252>

Application Note