Software Application Profile: TriplotGUI, a molecular epidemiology toolbox for investigating associations between exposures, omics, and

International Journal of Epidemiology, 2025, **54(6)**, dyaf203
https://doi.org/10.1093/ije/dyaf203
Software Application Profile

Software Application Profile

Software Application Profile: TriplotGUI, a molecular epidemiology toolbox for investigating associations between exposures, omics, and outcomes

Yingxiao Yan[1,*], Anton Ribbenstedt[1], Tessa Schillemans[2,3], Carl Brunius[1]

[1]Food and Nutrition Sciences, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden
[2]Division of Family Medicine, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden
[3]Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

*Corresponding author. Department of Life Sciences, Chalmers University of Technology, Chalmersplatsen 4, Gothenburg, SE-412 96, Sweden.
E-mail: yingxiao@chalmers.se

Keywords: molecular epidemiology; Shiny application; meet-in-the-middle analysis; mediation analysis; omics.

Key Features

- TriplotGUI is a user-friendly molecular epidemiology toolbox for analyzing, visualizing, and interpreting associations between exposures, Omics data, and outcomes.
- TriplotGUI provides streamlined workflow and advanced visualizations for meet-in-the-middle and mediation analysis.
- TriplotGUI is freely accessible both online and as a standalone application.

Introduction

Observational epidemiological research has evolved from single exposure–outcome studies to exposure-wide and outcome-wide investigations [1, 2]. Omics technologies have further enabled molecular epidemiology to mechanistically link exposures to outcomes via molecular data as potential mediators [3]. By providing comprehensive views of biological systems across genomics, transcriptomics, proteomics, and metabolomics, omics technologies capture the cumulative effects of environmental and endogenous factors over time [4], helping to elucidate the biochemical mechanisms underlying disease pathophysiology [5]. For example, metabolite profiling, representing a "snapshot" of the human metabolome, is frequently used to explore exposome-metabolome-disease linkages [6].

However, omics data typically contain a large number of molecular features [7] and mechanistic investigations using omics necessitate the selection of exposure- and outcome-relevant features. Common approaches include variable selection through supervised machine learning or performing univariate analysis, which allows confounder adjustment and more straightforward interpretation. However, univariate analysis struggles with collinearity, potentially misattributing effects among correlated variables and distorting their individual contributions, increasing false discovery rates [8]. Supervised machine learning can both select predictive features and manage collinearity [8]. However, it offers limited interpretability and confounder adjustment is often not possible [9]. Moreover, both methods rely on arbitrary thresholds (e.g. *P* values or importance rankings), introducing subjectivity in feature selection.

To address these issues, integrating machine learning with linear models seems intuitively appealing, potentially mitigating their individual limitations while enhancing feature selection robustness and interpretability. Building on this premise, we developed the triplot R package [10], which employs principal component analysis on omics variables, followed by a meet-in-the-middle approach [11] using linear methods and the co-visualization of the components' associations with both exposures and outcomes, adjusting for confounders. The triplot enables the exploration of biological mechanisms from exposures to outcomes through molecular features that dominate the components-of-interest, avoiding strict variable selection thresholds [10]. Furthermore, a triplot facilitates holistic investigations of multi-exposure and multi-outcome relationships via omics mediators.

While the triplot algorithm advanced holistic analysis in molecular epidemiology, it also has several limitations: it employs meet-in-the-middle for integrative analysis, but lacks formal mediation analysis [12], constraining the causal interpretation of exposure–mediator–outcome pathways. Moreover, the triplot did not support categorical variables with more than two classes, limiting analyses involving polytomous exposures and outcomes [e.g. smoking status and body mass index (BMI) categories]. Additionally, omics data reduction was limited to

Received: 12 March 2025; **Accepted:** 30 October 2025

principal component analysis (PCA). Finally, the package required users to possess R programming skills for its non-automated analysis steps.

To overcome these challenges, we developed the TriplotGUI tool—a user-friendly interface that streamlines data analysis and the visualization of exposures, omics, and outcomes. TriplotGUI implements the following steps and functionalities:

- Step 1: Data reduction of omics variables. Omics variables are transformed into components by using either PCA or weighted correlation network analysis (WGCNA) [13].
- Step 2: Associations with exposures. Associations are calculated between exposures and component scores, potentially adjusting for confounders. For continuous exposures, Pearson, Spearman, or Kendall correlations are used, whereas linear models are applied for categorical exposures.
- Step 3: Associations with outcomes. Risk associations between outcomes and component scores are assessed, potentially adjusting for confounders. Generalized linear models are used for continuous and binary outcomes. For categorical outcomes (more than two classes), either one-hot encoding with generalized linear models or multinomial models are applied. If pairing information is provided (e.g. matched case–control pairs), then linear mixed models are used for continuous outcomes and conditional logistic regression is used for binary outcomes.
- Step 4: Meet-in-the-middle triplot visualization. Co-visualization of omics components (loadings and optionally scores), exposure association estimates (correlation coefficients for continuous variables and beta-coefficients for categorical variables), and outcome risk-association estimates (beta-coefficients or odds ratios) as layers in a 2D plot.
- Step 5: Mediation analysis and visualization. Mediation analysis is performed on user-specified exposures and outcomes, using either component scores or individual omics variables as mediators. Both conventional (product method) [12] and counterfactual [14] approaches are supported, with flexible confounder adjustment in exposure–mediator and mediator–outcome models. Direct and indirect effects are visualized in a barplot, showing their direction and magnitude.

TriplotGUI also provides:

- Comparative visualization: Exposure and risk-association estimates, mediation estimates, proportion mediated, and *P* values (star system) are visualized in heat maps, facilitating component selection.
- Data download: Intermediate data and generated results including effect estimates and *P* values can be inspected and downloaded.

A detailed tutorial is available at https://metabocomp.github.io/TriplotGUI_tutorial.

Implementation

TriplotGUI is available both as an online tool at https://metabocomp.shinyapps.io/triplotgui and as a standalone application installable from https://github.com/MetaboComp/TriplotGUI. For sensitive datasets, we recommend the locally installed standalone version, preferably in a safe environment to ensure that data remain secure and are not transmitted externally.

TriplotGUI was developed in R using the Shiny package, enabling interactive analysis pipelines through a graphical user interface. For standalone use, after installing the package and dependencies, users can run TriplotGUI::TriplotGUI_shiny() to launch the application. Detailed setup guidance is provided in the online tutorial.

The application's reactive stepwise modular design organizes workflows into five analysis steps and two additional functionalities, enhancing upgradability and pipeline transparency. Fig. 1a (flow chart) and Fig. 1b (sidebar layout) illustrate how the interface modules correspond to the analytical steps and functionalities. Further details on TriplotGUI's modular workflow and technical specifications for data handling are provided in Supplementary Text S1.

Use

This section demonstrates TriplotGUI usage through a practical example. The dataset, available at https://github.com/MetaboComp/TriplotGUI/tree/main/Example_data/Example2, consists of simulated data derived from the Swedish Västerbotten Intervention Programme cohort, with processing methodology detailed in Schillemans *et al.* [10]. The dataset includes 1000 observations (500 matched case–control pairs by gender and age), featuring 17 dietary exposure variables (dietary indices and food intake), 31 omics variables, and 3 outcomes: binary type 2 diabetes (T2D), continuous BMI, and categorical BMI (four levels: obese, overweight, normal, underweight). Five potential confounders serve as covariates (gender, age, education, physical activity, smoking), alongside auxiliary data including pairing information matching controls to cases. The complete dataset contains no missing values and has consistent observation ordering across dataframes. This demonstration explores exposure–outcome associations via metabolites as molecular mediators.

The TriplotGUI workflow begins with omics data upload (Step 1), in which dimensionality reduction via PCA or WGCNA produces component scores and loadings. Exposure and outcome data are uploaded in Steps 2 and 3, with optional simultaneous covariate upload for confounding adjustment. Importantly, TriplotGUI permits data inspection, variable class modification, and the removal of undesirable variables for each dataset. For this use case, all metabolomics and exposure variables are converted into numeric form, with auxiliary or covariate data excluded. Advanced applications are detailed in the online tutorial.

PCA implementation in Step 1 generates five principal components by default (user-adjustable). Steps 2 and 3 quantify component associations with exposures/outcomes, visualized as heat maps through the "Comparative Visualization" module (Supplementary Fig. S1). Noticeably, Component 1 showed strong correlation with multiple exposures and Component 2 exhibited the strongest association with BMI and was therefore used to generate the triplot in Step 4. However, both Components 3 and 4 demonstrated potentially interesting associations, warranting further investigation.

In Step 4, the triplot co-visualizes omics variable loadings (black arrows), correlation coefficients (blue circles), and risk estimates (red squares), providing a global view of exposure–omics–outcome associations (Supplementary Fig. S2a).

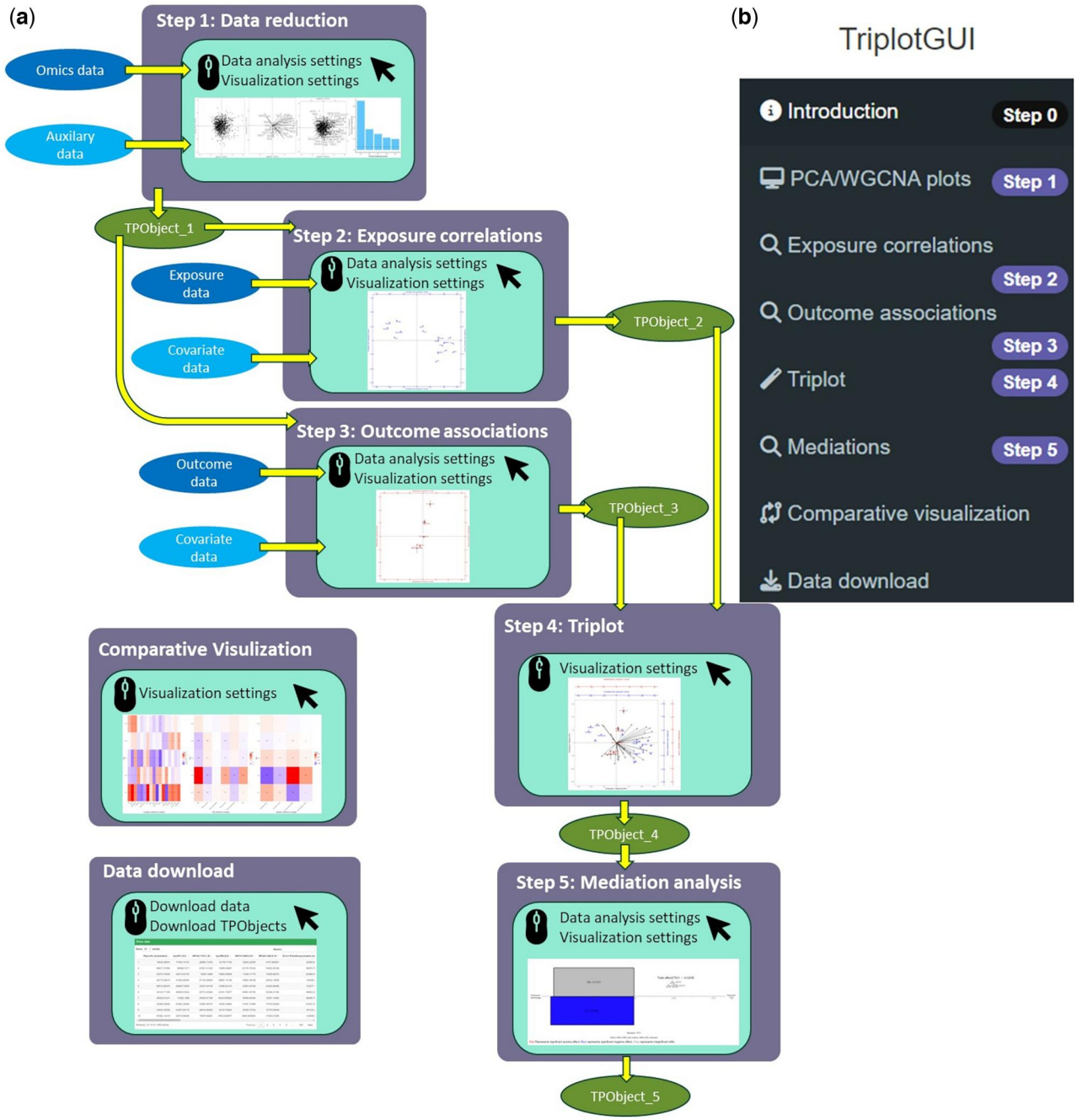

Figure 1. (a) Flow chart of the TriplotGUI structure. Each box corresponds to a tab in the TriplotGUI sidebar, representing one of the five main steps or two additional functionalities. Arrows connecting the boxes can be interpreted as passing data. Ovals represent different types of input and output: essential input data; covariate and auxiliary input data; and TriplotGUI objects ("TPObjects") generated in the steps, which contain the accumulated information gathered up to and including that step. (b) Sidebar from the TriplotGUI application that corresponds to the steps and functionalities visualized in the flow chart.

Compared with Component 2, Component 1 shows overall stronger correlations with food items and dietary indexes, but weaker associations with BMI or T2D, suggesting a dietary-influenced metabolite pattern with limited metabolic health implications. Conversely, Component 2 is more strongly associated with BMI and T2D, and also with certain unhealthy exposures (e.g. sausage), possibly indicating how certain food choices can affect health, potentially mediated through metabolite patterns represented as loadings in the two visualized components. In addition to data reduction using PCA in Step 1, users can also opt to use WGCNA for corresponding triplots (Supplementary Text S2 and Supplementary Fig. S2b).

Based on the "Comparative Visualization" and triplot from Step 4, users can select specific exposures and outcomes for mediation analysis in Step 5, facilitating further investigation of the potential biological mechanisms underlying the components-of-interest.

In the current use case on BMI and T2D, using PCA for omics data reduction, we focus on molecular mediation for the Hamburger and Baltic Sea Dietary Score—the dietary exposures that showed the strongest associations with the

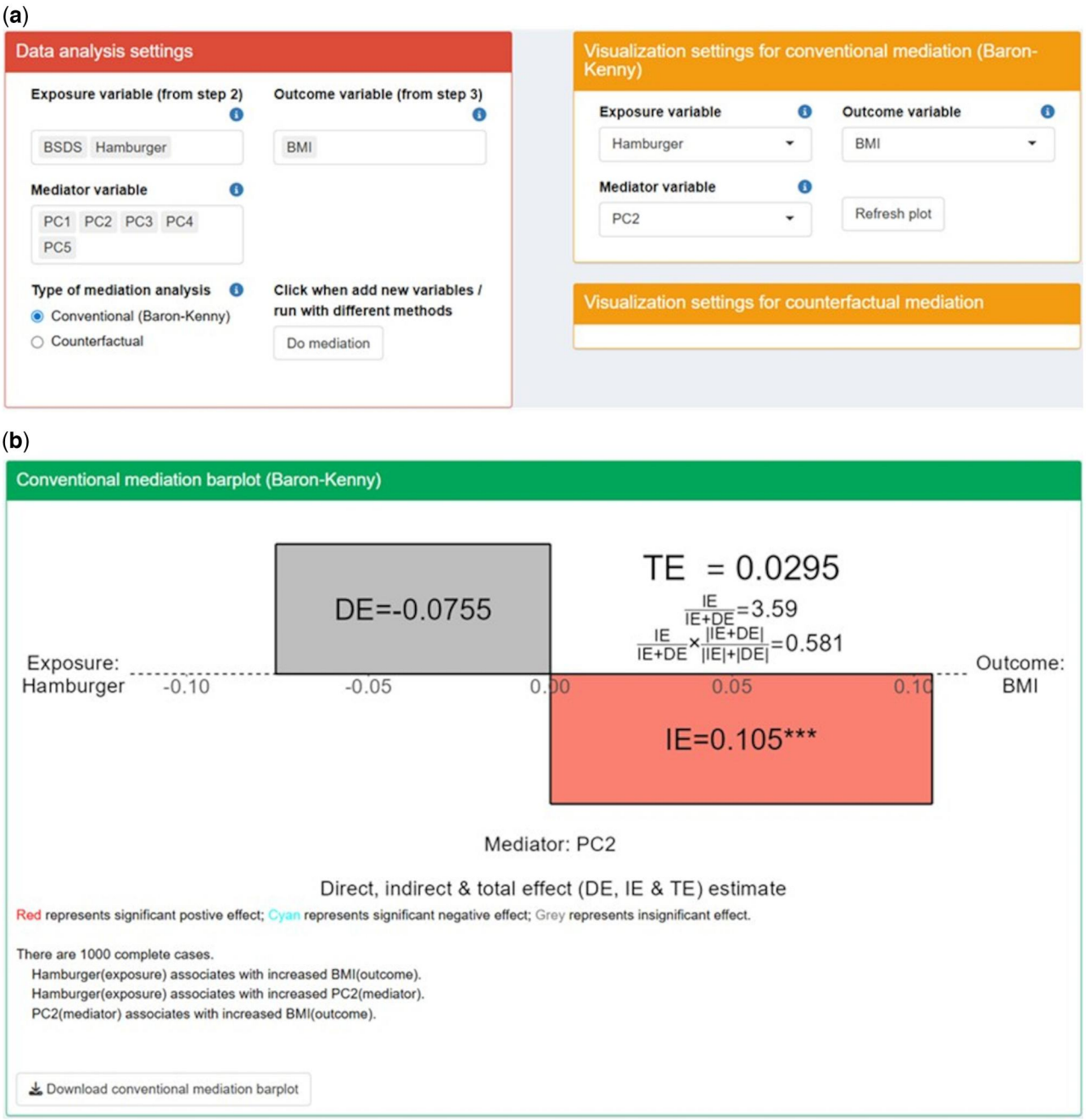

Figure 2. Mediation analysis in TriplotGUI Step 5. (a) Data-analysis and visualization settings. Mediation analysis using the Baron–Kenny method is performed for the selected exposures and outcomes. Specific exposure–component–outcome combinations are selected for visualization. (b) Mediation barplot using hamburger consumption as the exposure, Component 2 as the mediator, and BMI as the outcome. The positive indirect effect is consistent with having similar directions between the hamburger–component correlation and the component–BMI association. In contrast, direct and total effects were not observed (P value $> .05$). Positive and negative effects are distinguished by their position relative to the axis. IE, indirect effect; DE, direct effect; TE, total effect. *P value $< .05$, **P value $< .01$, ***P value $< .001$.

omics components. Here, we demonstrate conventional mediation analysis by using the Baron–Kenny (product) method [12] (Fig. 2), although TriplotGUI also offers counterfactual mediation analysis via the "mediation" package [14] (Fig. 2a). Step 5 visualizes the mediation estimates (i.e. direct, indirect, total effect, and proportion mediated) and association directions for selected exposure–component–outcome mediations (Fig. 2b). As the requirement for an apparent total effect to justify mediation analysis has been largely abandoned in modern statistical practice [15, 16], we therefore focus directly on the indirect effects for its potential to reveal mechanistic insights. Both mediation and meet-in-the-middle analyses consistently support the indirect effects of hamburger consumption (a marker of an unhealthy diet) on BMI. Specifically, Component 2 was associated with higher BMI and hamburger intake in the meet-in-the-middle analysis (Supplementary Fig. S2) and showed a corresponding positive indirect effect in the mediation analysis (Fig. 2b). This

suggests that the metabolites represented by Component 2 may mediate an association of hamburger consumption with increased BMI, although a total effect could not be observed (P value = .55). In addition, an unobserved direct effect (P value = .07) implies that other molecular mechanisms are either absent or sum up to null.

Notably, the "Comparative Visualization" automatically updates with mediation analysis results from Step 5 (Supplementary Fig. S3). All data generated throughout the analyses can be downloaded in the "Data Download" module, enabling further exploration and refined analyses.

This use case highlights how TriplotGUI's user-friendly interface streamlines holistic analysis and the visualization of molecular mediation between exposures and outcomes, lowering the technical barriers for users with limited coding experience. Omics features dominating the components-of-interest in TriplotGUI can be further investigated for biological mechanisms linking specific exposures to outcomes. While the use case illustrates a simplified TriplotGUI workflow, detailed guidance on advanced functionalities, including adjustment for confounders at various stages and counterfactual mediation analysis, is available in the "Manual" section of the web tutorial.

Discussion

With its automated workflow and advanced visualizations, TriplotGUI supports researchers in analysing complex relationships among exposure(s), omics, and outcome(s). The tool accelerates analysis for users with limited R knowledge and enables comprehensive exploratory investigations through its diverse functionalities. While the interface is user-friendly, advanced users can also integrate custom functions, offering flexibility to adapt the package for more complex analyses and causal structures.

Beyond data reduction, exposure–outcome associations, and mediation analysis, TriplotGUI provides flexible options for handling categorical variables. Users can choose between multinomial regression and one-hot encoding with subsequent logistic regressions for categorical outcome variables (more than two classes). For binary matched case–control outcomes, conditional logistic regression is available by using pairing information uploaded as auxiliary data.

While TriplotGUI automates data analysis and visualization, we acknowledge several limitations: TriplotGUI assumes that omics variables represent potential molecular mediators between exposures and outcomes. However, causal relationships cannot be tested within the tool itself and users must assess the suitability for their research questions. Moreover, continuous exposures are investigated by using correlation analysis, corresponding to a cross-sectional design. Linear models reflecting temporality may be more appropriate if exposure data precede omics assessment. In addition, TriplotGUI currently supports confounder adjustment for exposure–outcome associations via molecular mediators, but not more complex confounding issues [17] and effect modification [18]. Advanced users, however, can replace the built-in mediation procedures with custom scripts while leveraging TriplotGUI's workflow and visualizations. Furthermore, survival analysis [19], linear mixed-effect models for repeated measurement [20] and regularization techniques [21] have not yet been incorporated for risk modeling. Moreover, TriplotGUI's mediation analysis is currently restricted to single mediators and it may benefit from integrating advanced high-dimensional mediation techniques, accounting for multiple mediators and enabling simultaneous mediator selection [17–19]. Additional methodological discussions on mediation analysis are provided in Supplementary Text S3. TriplotGUI's modular design allows future upgrades to address current limitations, with plans to expand data reduction, risk modeling, and mediation analysis [18, 22] as methodologies evolve.

While TriplotGUI's programming-free interface improves accessibility, users still need to have domain-specific epidemiological knowledge to frame the research questions, define the exposures/outcomes, construct causal diagrams, and interpret the results. Data preprocessing (e.g. imputing missing values, normalization) must be handled externally due to the variety of options. Though tested primarily with metabolomics, TriplotGUI supports other omics, such as proteomics. However, TriplotGUI faces computational limits with large-scale transcriptomic/epigenomic data. We therefore suggest the preselection of variables of interest from extensive predictor data prior to analysis with TriplotGUI by using variable selection techniques designed to minimize data leakage, such as the MUVR2 approach [23]. In addition, future updates will optimize scalability via parallel computing and memory management. To further enhance usability and interpretability, interactive visualizations with dynamic inspection of results (e.g. zooming, tooltips) are also planned for future versions.

In summary, TriplotGUI streamlines the investigation of exposure–omics–outcome relationships by reducing technical barriers, automating analyses, and providing intuitive visualizations. As a freely available application, TriplotGUI is designed to be maintained and upgraded to meet evolving research needs in the coming years.

Acknowledgements

We would like to thank Dr Claudio Cappadona from the Department of Pediatrics of Erasmus MC, University Medical Center Rotterdam (Rotterdam, The Netherlands) for his valuable insights and suggestions that helped in improving our R package.

Author contributions

Creation of the Shiny app: Y.Y. Data analysis: Y.Y. Study concept and design: C.B. and Y.Y. Manuscript writing: Y.Y. and C.B. Project supervision: C.B. and A.R. Interpretation of the data and critical revision and editing of the manuscript: all authors. All authors read and approved of the final manuscript.

Supplementary data

Supplementary data is available at *IJE* online.

Conflict of interest

None declared.

Funding

This work was supported by a Swedish Research Council Formas grant (2020–01653) and funding from European

Union's Horizon 2020 research and innovation programme under grant agreement no. 874739.

Data availability

TriplotGUI is accessible online at https://metabocomp.shinyapps.io/triplotgui. The TriplotGUI source code and test data are freely available on Gitlab at https://github.com/MetaboComp/TriplotGUI. A comprehensive tutorial is available at https://metabocomp.github.io/TriplotGUI_tutorial.

Use of artificial intelligence (AI) tools

The authors utilized Perplexity Pro for text editing during the manuscript preparation and take full responsibility for the accuracy and integrity of the published content.

References

1. Ioannidis JPA. Exposure-wide epidemiology: revisiting Bradford Hill. *Stat Med* 2016;**35**:1749–62.
2. Vanderweele TJ. Outcome-wide epidemiology. *Epidemiology* 2017;**28**:399–402.
3. Schneider MV, Orchard S. Omics technologies, data and bioinformatics principles. 2011. https://doi.org/10.1007/978-1-61779-027-0_1
4. Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 2012;**13**:263–9.
5. Walker DI, Valvi D, Rothman N *et al.* The metabolome: a key measure for exposome research in epidemiology. *Curr Epidemiol Rep* 2019;**6**:93–103.
6. Cheng S, Larson MG, McCabe EL *et al.* Distinct metabolomic signatures are associated with longevity in humans. *Nat Commun* 2015;**6**:7317. https://doi.org/10.1038/ncomms7791
7. Feldner-Busztin D, Nisantzis PF, Edmunds SJ *et al.* Dealing with dimensionality: the application of machine learning to multi-omics data. *Bioinformatics* 2023;**39**:btad021. https://doi.org/10.1093/bioinformatics/btad021
8. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol* 2015;**3**:23.
9. Posma JM, Garcia-Perez I, Ebbels TMD *et al.* Optimized phenotypic biomarker discovery and confounder elimination via covariate-adjusted projection to latent structures from metabolic spectroscopy data. *J Proteome Res* 2018;**17**:1586–95.
10. Schillemans T, Shi L, Liu X *et al.* Visualization and interpretation of multivariate associations with disease risk markers and disease risk—the triplot. *Metabolites* 2019;**9**:133.
11. Chadeau-Hyam M, Athersuch TJ, Keun HC *et al.* Meeting-in-the-middle using metabolic profiling-a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers* 2011;**16**:83–8.
12. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research. conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;**51**:1173–82.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559. https://doi.org/10.1186/1471-2105-9-559
14. Tingley D, Yamamoto T, Hirose K *et al.* mediation: R package for causal mediation analysis. *J Stat Soft* 2014;**59**:
15. Hayes AF. Beyond Baron and Kenny: statistical mediation analysis in the new millennium. *Commun Monogr* 2009;**76**:408–20.
16. Zhao X, Lynch JG, Chen Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. *J Consum Res* 2010;**37**:197–206.
17. Clark-Boucher D, Zhou X, Du J *et al.* Methods for mediation analysis with highdimensional DNA methylation data: Possible choices and comparisons. *PLoS Genet* 2023;**19**:e1011022.
18. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiol Methods* 2013;**2**:95–115.
19. Dai JY, Stanford JL, LeBlanc M. A multiple-testing procedure for high-dimensional mediation hypotheses. *J Am Stat Assoc* 2022;**117**:198–213.
20. Bolker BM, Brooks ME, Clark CJ *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 2009;**24**:127–35. https://doi.org/10.1016/j.tree.2008.10.008
21. De Mol C, De Vito E, Rosasco L. Elastic-net regularization in learning theory. *J Complex* 2009;**25**:201–30. https://doi.org/10.1016/j.jco.2009.01.002
22. Xu Z, Li C, Chi S *et al.* Speeding up interval estimation for R2-based mediation effect of high-dimensional mediators via cross-fitting. *Biostatistics* 2024;**25**:kxae037. https://doi.org/10.1093/biostatistics/kxae037.
23. Yan Y, Schillemans T, Skantze V *et al.* Adjusting for covariates and assessing modeling fitness in machine learning using MUVR2. *Bioinform Adv* 2024;**4**:vbae051. https://doi.org/10.1093/bioadv/vbae051

International Journal of Epidemiology, 2025, 54, 1–6
https://doi.org/10.1093/ije/dyaf203
Software Application Profile