

¹ Taming your metabolic datasets with MeDUSA

² **Laura Ann Castaneda**  ^{1*}, **Eric Hetzel** ^{2*}, and **Ahmed Ali**  ^{1¶}

³ **1** Leiden University, Leiden, Netherlands **2** Independent Researcher, Netherlands ¶ Corresponding author
⁴ * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright
¹⁵ and release the work under a
¹⁶ Creative Commons Attribution 4.0
¹⁷ International License ([CC BY 4.0](#)).
¹⁸

⁵ Summary

⁶ Over ten trillion cells are hard at work in the human body([Bianconi, 2013](#)) and there can be
⁷ significant heterogeneity amongst them affecting biological development, disease progression,
⁸ and treatment response([Zhang & Vertes, 2018](#)).

⁹ One technique to categorize this heterogeneity is single cell analysis using mass spectrometry.
¹⁰ However this technique introduces new challenges. One of which is the small sample volume,
¹¹ which limits the possibilities for separation and thus makes the data non-compatible with
¹² traditional analysis pipelines. We introduce the R package MeDUSA for Metabolomic Direct-
¹³ infusion Untargeted Single-cell Analysis.

¹⁴ MeDUSA is a start-to-finish analysis package allowing metabolomics researchers to focus on
¹⁵ analytical content rather than R proficiency. MeDUSA handles the suggested workflow by Southam
¹⁶ ([Southam et al., 2017](#)) from data extraction to filtering without a chromatogram, carrying the
¹⁷ data through statistical analysis and identification of features for biological interpretation.

¹⁸ Statement of Need

¹⁹ Due to the small volume of a single cell, direct infusion(DI), nano-electrospray ionization
²⁰ (nESI) is a highly suitable technique. However liquid chromatography mass spectrometry
²¹ (LC-MS) is significantly more common than DI mass spectrometry, therefore most software
²² is being developed for separation chromatography. For instance, XCMS offers filtering and
²³ statistical analysis with visualization, however, all of the filtering is reliant on the presence
²⁴ of a chromatogram([Smith et al., 2006](#)). Similarly, MetaboAnalyst has an impressive inter-
²⁵ face that allows for metabolomic pathway analysis, but the full potential is only unlocked
²⁶ with a chromatogram([Pang, 2021](#)). Even paid subscription based software such as Thermo
²⁷ Fisher Scientific's Compound Discoverer is designed to align and filter peaks based on a
²⁸ chromatogram([Cooper & Yang, 2024](#)). Furthermore, the mentioned software does not allow
²⁹ the user to define preprocessing methods such as centroiding or alignment. In contrast, the
³⁰ modularity of MeDUSA is built specifically for direct infusion data, and it lays the framework
³¹ for method specification. The modularity also offers the user the ability to bypass functions,
³² introduce external functions to filter, and process the data as they see fit, an option that other
³³ software does not offer. Currently, if researchers want this level of modularity and customization
³⁴ designed for single cell data, they must write their own scripts which requires proficiency
³⁵ in a programming language and a significant time allotment. Therefore the metabolomic
³⁶ community needs a software option that will enable complete, modular, and customizable
³⁷ processing of mass spectra without a chromatogram.

38 Description

39 The goal of MeDUSA is to provide a toolset that is modular, customizable, and user friendly.
 40 There are five major sections along this standard flow as shown in [Figure 1](#).
 41 Modularity is achieved by using standardized interoperable data-objects. This allows the user
 42 to choose any collection and order of functions as they see fit. See the README for a list
 43 and description of the standard objects (<https://github.com/laura-hetzel/MeDUSA#data-structures>).
 44
 45 Customization is achieved by being greatly parameterized and leveraging modularity. This
 46 allows the user to dial in their variables, such as thresholds, aggregation methods, tolerances,
 47 etc. The user may also interrupt the suggested flow to perform any custom logic to their
 48 needs, and reintroduce their updated data into the MeDUSA flow, so long as the data structure
 49 is maintained. The three primary data-objects have the same structure and are differentiated
 50 by name for human readability along the suggested flow. However, they are technically
 51 interchangeable, thus increasing customization.
 52 User-friendliness is achieved with “magic” functions, readability, and containerization. The
 53 magic functions leverage suggested parameter values and simplify many functions within each
 54 of the five major sections. These methods can allow a user to go from mzML files to a list of
 55 compounds in few commands. [Figure 2](#) illustrates a detailed list of functions, and how the
 56 magic functions simplify them. To ease readability, files and methods are prepended with the
 57 expected input data-object type. This naming convention helps users identify what methods
 58 are available to them at different stages of the suggested flow. Containerization manages
 59 dependencies and provides HMDB and lipid data for convenient m/z to compound mapping.

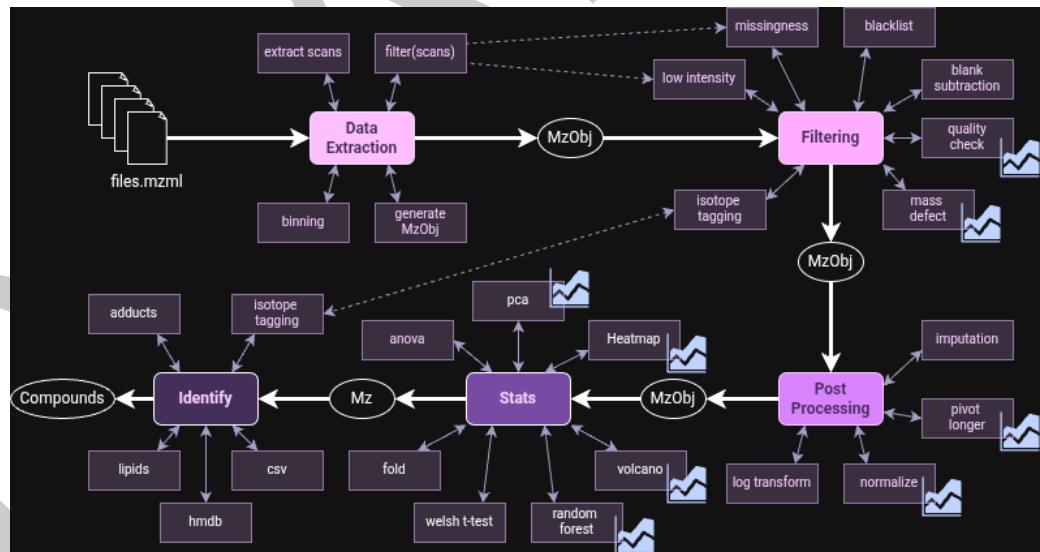


Figure 1: Map of the five sections of MeDUSA and the capabilities of each section. The bold arrows indicate a suggested workflow. The dashed arrows indicate references. The circled text indicated the object data type. The plot symbol indicates the function may output a plot.

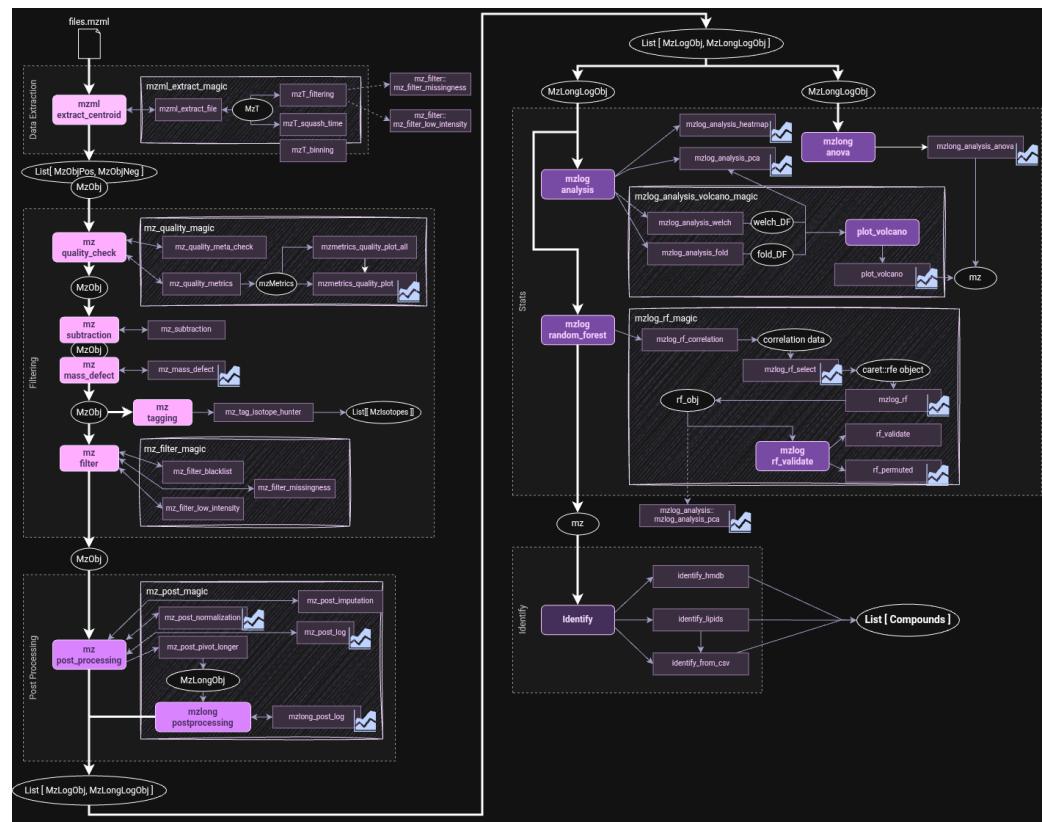


Figure 2: Detailed suggested flow and function map. Note the “magic” functions which aggregate similar functions for user ease. The bold arrows indicate a suggested workflow. The circled text indicated the object data type. The plot symbol indicates the function may output a plot.

60 Research projects using the software

61 Current research projects relating to single cell metabolic profiling of the cell cycle of Fucci
 62 cells, stem cell differentiation, hypoxic organoids, and metastatic organoids would benefit from
 63 this package. The projects currently use two different commercially available software as well
 64 as R scripts that lack robustness for the data analysis; shifting the analysis to MeDUSA will
 65 enable timely and reliable results. Upon the release of the package, it will be implemented lab
 66 wide for live single cell metabolomics.

67 Acknowledgements

68 This research was funded in part by the European Union’s Horizon 2020 research and innovation
 69 program under the Marie Skłodowska-Curie grant agreement number 861196.

70 Rosa (species = Canis Familiaris, breed = Catahoula Leopard) was essential to the mental
 71 well-being and creativity of the authors, and therefore essential to the project.

72 We would like to thank Alida Kindt and Thomas Hankemeier for their support and direction
 73 on this project.

74 References

- 75 Bianconi, E. et al. (2013). An estimation of the number of cells in the human body. *Ann.*
76 *Hum. Biol.*, 40. <https://doi.org/10.3109/03014460.2013.807878>
- 77 Cooper, B., & Yang, R. (2024). An assessment of AcquireX and Compound Discoverer
78 software 3.3 for non-targets metabolomics. *Sci. Rep.*, 14. <https://doi.org/10.1038/s41598-024-55356-3>
- 80 Pang, Z. et. al. (2021). MetaboAnalyst 5.0: narrowingthe gap between raw spectra and
81 functional insights. *Nucleic Acids Res.*, 49. <https://doi.org/10.1093/nar/gkab382>
- 82 Smith, C. A., Want, E. J., O'Maille, G., & Siuzdak, G. (2006). XCMS: Processing Mass
83 Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching,
84 and Identification. *Anal. Chem.*, 78. <https://doi.org/10.1021/ac051437y>
- 85 Southam, A. D., Weber, R. J. M., Engel, J., Jones, M. R., & Viant, M. R. (2017). A
86 complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion
87 mass-spectrometry-based metabolomics and lipidomics. *Nat. Protoc.*, 12. <https://doi.org/10.1038/nprot.2016.156>
- 89 Zhang, L., & Vertes, A. (2018). Single-Cell Mass Spectrometry Approaches to Explore Cellular
90 Heterogeneity. *Angew. Chem. Int.*, 57. <https://doi.org/10.1002/anie.201709719>

DRAFT