

1 MeDUSA: Taming your metabolic datasets using an 2 R-package for data processing of Direct Infusion 3 Untargeted Single-cell Analysis

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

5 ¹ Leiden University, Leiden, Netherlands ² Independent Researcher, Netherlands ¶ Corresponding author
6 * 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Published: unpublished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

17 Authors of papers retain copyright
18 and release the work under a
19 Creative Commons Attribution 4.0
International License ([CC BY 4.0](#)).
20

Summary

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

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

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

Statement of Need

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

40 Description

41 The goal of MeDUSA is to provide a toolset that is modular, customizable, and user-friendly.
 42 There are five major sections along this standard flow as shown in [Figure 1](#).

43 Modularity is achieved by using standardized interoperable data objects. This allows the user
 44 to choose any collection and order of functions that they see fit. See the README for a list
 45 and description of the standard objects (<https://github.com/laura-hetzel/MeDUSA#data-structures>).

46 Customization is achieved by being greatly parameterized and leveraging modularity. This
 47 allows the user to dial in their variables, such as thresholds, aggregation methods, tolerances,
 48 etc. The user may also interrupt the suggested flow to perform any custom logic according
 49 to their needs, and reintroduce their updated data into the MeDUSA flow, so long as the data
 50 structure is maintained. The three primary data objects have the same structure and are
 51 differentiated by name for human readability along the suggested flow. However, they are
 52 technically interchangeable, thus increasing customization.

53 User-friendliness is achieved with “magic” functions, readability, and containerization. The
 54 magic functions leverage suggested parameter values and simplify many functions within each
 55 of the five major sections. These methods can allow a user to go from mzML files to a list of
 56 compounds in few commands. [Figure 2](#) illustrates a detailed list of functions, and how the
 57 magic functions simplify them. To ease readability, files and methods are prepended with the
 58 expected input data-object type. This naming convention helps users identify what methods
 59 are available to them at different stages of the suggested flow. Containerization manages
 60 dependencies and provides HMDB and lipid data for convenient m/z to compound mapping.

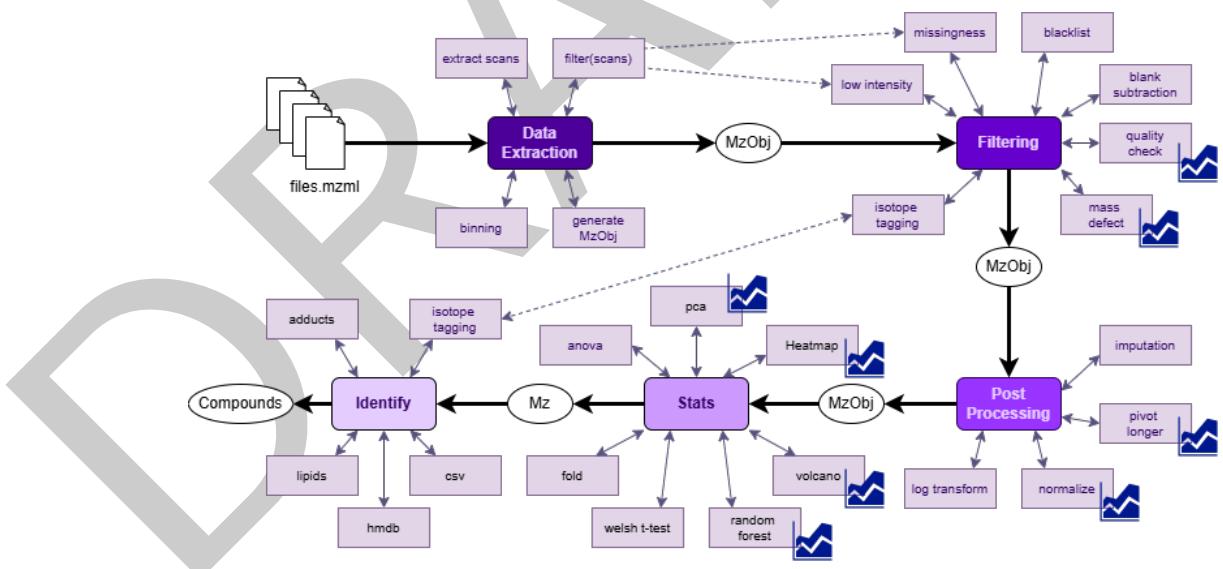


Figure 1: Map of the five sections of MeDUSA and the capabilities of each section. The bold arrows indicate a suggested workflow. The dashed arrows indicate references. The circled text indicated the object data type. The plot symbol indicates the function may output a plot.

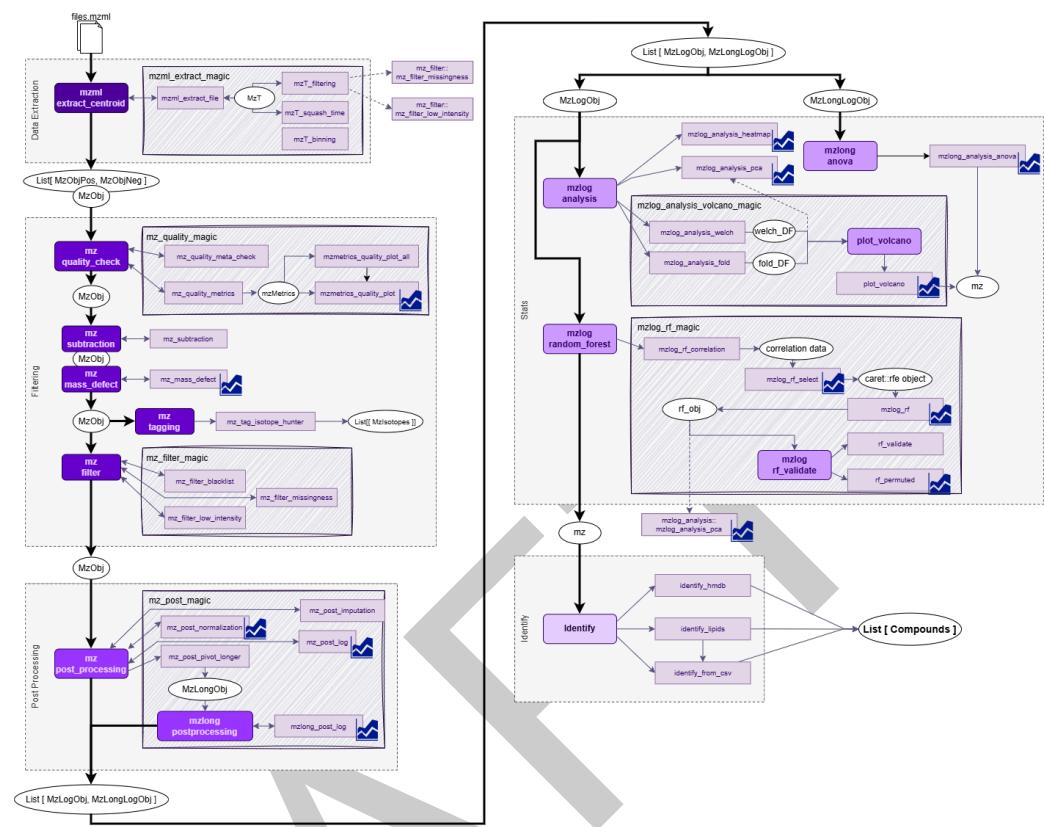


Figure 2: Detailed suggested flow and function map. Note the “magic” functions which aggregate similar functions for user ease. The bold arrows indicate a suggested workflow. The circled text indicated the object data type. The plot symbol indicates the function may output a plot.

62 Research projects using the software

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

69 Acknowledgements

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

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

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

76 References

- 77 Bianconi, E. et al. (2013). An estimation of the number of cells in the human body. *Ann.*
78 *Hum. Biol.*, 40. <https://doi.org/10.3109/03014460.2013.807878>
- 79 Cooper, B., & Yang, R. (2024). An assessment of AcquireX and Compound Discoverer
80 software 3.3 for non-targets metabolomics. *Sci. Rep.*, 14. <https://doi.org/10.1038/s41598-024-55356-3>
- 82 Pang, Z. et al. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and
83 functional insights. *Nucleic Acids Res.*, 49. <https://doi.org/10.1093/nar/gkab382>
- 84 Smith, C. A., Want, E. J., O'Maille, G., & Siuzdak, G. (2006). XCMS: Processing mass
85 spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and
86 identification. *Anal. Chem.*, 78. <https://doi.org/10.1021/ac051437y>
- 87 Southam, A. D., Weber, R. J. M., Engel, J., Jones, M. R., & Viant, M. R. (2017). A
88 complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion
89 mass-spectrometry-based metabolomics and lipidomics. *Nat. Protoc.*, 12. <https://doi.org/10.1038/nprot.2016.156>
- 91 Zhang, L., & Vertes, A. (2018). Single-cell mass spectrometry approaches to explore cellular
92 heterogeneity. *Angew. Chem. Int.*, 57. <https://doi.org/10.1002/anie.201709719>

DRAFT