Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Information 2: Instructions for using NuMo Finder

Monitoring nucleoside modifications are critical to the general area of epigenetic and epitranscriptomic research, as a large majority of DNAs and RNAs are modified. NuMoFinder is an open-source software tool developed using Python and the MatchMS library.[1] It's designed to automatically identify and quantify the peaks of nucleic acid modifications referred from Modomics and DNAmod databases across a series of mass spectrometry files using the matchMS package.[2, 3] Meanwhile, NuMo Finder incorporates the unknown modification (untargeted) search relying on the major ribose/deoxyribose lost product ions of nucleosides generated from collision-induced dissociation (CID). To get started with NuMoFinder, you'll need to input data in .mzXML format. The modifications recorded from Modomics and DNAmod databases have already been incorporated into the software, while the "CustomizedMods.csv" provides the searching of user-defined databases if needed.

The example data and results are available at: https://github.com/ChenfengZhao/NuMoFinder

Analysis setup:

[The configuration file named **config.ini** to control the behaviors of the software.]

input_path: folder contains MS data files (.mzXML)

output_path: folder to save the NuMo Finder search results (.mzXML)

customized_mods_list: the (optional) user-defined library that is included for searching

nucleoside_type: the type of analytes, DNA or RNA

permethyl: if permethylation is used for preparing the analytes

polarity: instrument polarity in MS analysis

ms1_mass_error_ppm: mass tolerance for MS1 searching

ms2_mass_error_ppm: mass tolerance for MS2 searching

min_rel_height: (optional) relative intensity threshold for searching

min_height: (optional) absolute intensity threshold for searching

min_mass: (optional) minimum *m/z* that is considered for searching

max_mass: (optional) maximum *m/z* that is considered for searching

gaussian_filter: (optional) Gaussian distribution is used for peak filtering

align_tolerance_min: the time tolerance for MS1 and MS2 peak alignment

unknown_search_mode: enable searching for unknown modification

min_height_unknow_search: the MS2 intensity threshold in the unknown search

flex_mode: (optional) ignoring of the requirements for monoisotopic distribution of analytes

Analysis Workflow:

Common Search Workflow: NuMo Finder initially reads and interprets the input data files, calculating the MS1 *m/z* values based on common modifications sourced from the Modomics and DNAmod databases. Additionally, a custom list feature lets users add their own modifications if they are not already present in the two aforementioned databases. If the computed masses fall within the mass range specified in config.ini, these masses are treated as MS1 *m/z* values and are searched for in the MS1 data. Subsequently, NuMo Finder determines ion distributions using the calculated MS1 masses and a preset MS1 mass error, in order to confirm the presence of these MS1 masses within a spectrum. By default, when a spectrum contains two isotopic masses that meet certain criteria, the software extracts details such as retention time, scan number, and peak intensity to generate RT-intensity curves for all the observed MS1 masses. To identify prominent peaks in the RT-intensity curve for the first MS1 mass, the software defaults to applying a Gaussian filter and logs all potential peaks, regardless of their absolute intensity. Similarly, for MS2 spectral analysis, the program checks for the presence of product mass in the MS2 spectrum. If this mass exceeds a predetermined minimum match count, relevant data like retention time, scan number, and peak intensity are logged to create an RT-intensity curve for each MS2 *m/z* value. Finally, NuMo Finder tallies the number of MS2 peaks that align with each MS1 peak by comparing their scan numbers or retention times.

Unknown Search Workflow: Initially, NuMo Finder identifies MS2 candidates based on various characteristics such as isotopic distribution, mass tolerance, and intensity. Ions already referenced in the database are considered known candidates and are therefore omitted from the

MS2 selection. Once the MS2 ion candidates are established, NuMo Finder computes the precursor mass by adding either the ribose mass (in the case of RNA) or deoxyribose mass (for DNA), subsequently creating an untargeted library that includes both MS1 precursor and MS2 product masses. This library serves as the input for a standard search workflow, employing the same search algorithm used in typical searches.

References

1. S. V. Florian Huber, Christiaan Meijer, Hanno Spreeuw, Efraín Manuel Villanueva Castilla, Cunliang Geng, Justin J. J. van der Hooft, Simon Rogers, Adam Belloum, Faruk Diblen, and Jurriaan H. Spaaks1, *Journal of Open Source Software*, 2020, **5**, 2411.
2. P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P. Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F. Aguilo and J. M. Bujnicki, *Nucleic Acids Research*, 2022, **50**, D231-D235.
3. A. J. Sood, C. Viner and M. M. Hoffman, *Journal of Cheminformatics*, 2019, **11**, 30.