Supporting Information

Autonomous METLIN-guided in-source fragment detection increases annotation confidence in untargeted metabolomics

Xavier Domingo-Almenara[1,*], J. Rafael Montenegro-Burke[1], Carlos Guijas[1], Erica L.-W. Majumder[1], H. Paul Benton[1], Gary Siuzdak[1,2,*]

[1] Scripps Center for Metabolomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States.

[2] Department of Molecular and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States.

* To whom correspondence should be addressed: xdomingo@scripps.edu, siuzdak@scripps.edu.

False discovery rate calculation:

We calculated the false discovery rate (FDR) for both the ratio score and the spectral similarity score (calculated as the dot product). The FDR was calculated using the following equation:

$$FDR = \frac{FP}{TP + FP} \tag{1}$$

Where FP and TP are the false positives and true positives, respectively. We calculated the FDR for different score segments, e.g., for 10%, 20%, ..., 80%. The lack of data above 80% precluded the calculation of the FDR. We calculated the FDR by dividing the TP identifications above each score value for segment, by the TP above the same score value. This lead to the FDR shown in Supplementary Figure 1. TP were calculated using the correctly identified metabolites. In some cases, for each correctly identified metabolite by MISA, false positive identifications were also reported. These false positive identifications were used as FP.

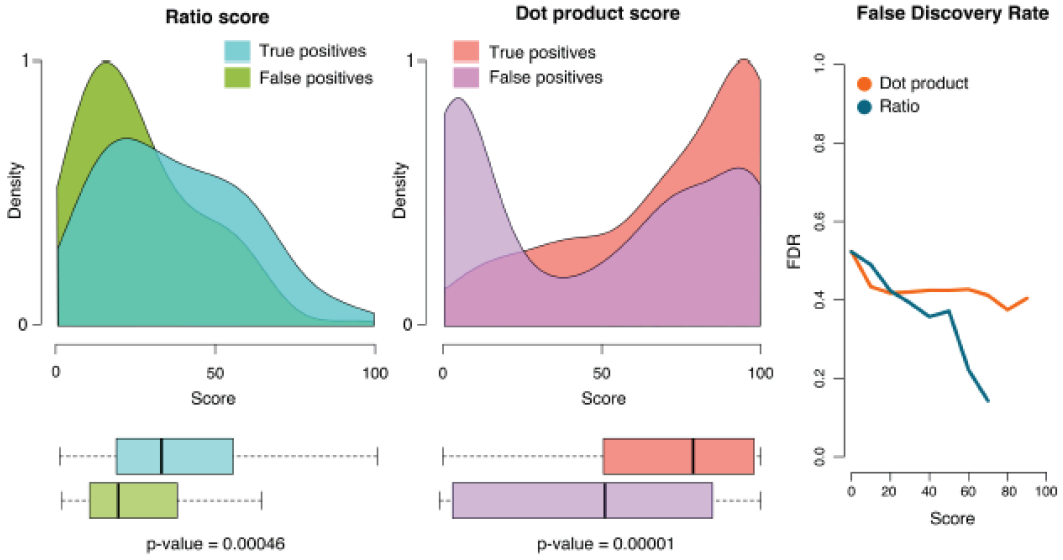

Figure 1: The density plots as well as the box plots of the TP and FP as a function of each score type is shown. The false discovery rate (FDR) is also shown. For this figure, the total number of FP and TP (at 0%) was of N=81 (TP) an N=90 (FP).