International Journal of Mass Spectrometry

Open search unveils modification patterns in formalin-fixed, paraffin-embedded thermo HCD and SCIEX TripleTOF shotgun proteomes

David L. Tabb [a b c d], Brandon D. Murugan [d], Javan Okendo [d], Omesan Naird [e], Jonathan M. Blackburn [d], Sindisiwe G.Buthelezi [f], Stoyan Stoychev [f]

[a] Bioinformatics Unit, South African Tuberculosis Bioinformatics Initiative, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Stellenbosch University, Cape Town, South Africa
[b] South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa
[c] Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa
[d] Institute of Infectious Disease and Molecular Medicine, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
[e] Department of Paediatrics, Red Cross Children's Hospital, University of Cape Town, Cape Town, South Africa
[f] Council for Scientific and Industrial Research, Pretoria, South Africa

https://doi.org/10.1016/j.ijms.2019.116266

Abstract

The application of database search algorithms with very wide precursor mass tolerances for the "Open Search" paradigm has brought new efforts at post-translational modification discovery in shotgun proteomes. This approach has motivated the acceleration of database search tools by incorporating fragment indexing features. In this report, we compare open searches and sequence tag searches of high-resolution tandem mass spectra to seek a common "palette" of modifications when analyzing multiple formalin-fixed, paraffin-embedded (FFPE) tissues from Thermo Q-Exactive and SCIEX TripleTOF instruments. While open search in MSFragger produced some gains in identified spectra, careful FDR control limited the best result to 24% more spectra than narrow search (worst result: a loss of 9%). Open pFind produced high apparent sensitivity for PSMs, but entrapment sequences hinted that the actual error rate may be higher than reported by the software. Combining sequence tagging, open search, and chemical knowledge, we converged on this set of PTMs for our four FFPE sets: mono- and di-methylation (nTerm and Lys), single and double oxidation (Met and Pro), and variable carbamidomethylation (nTerm and Cys).