thermoscientific

Estimating False Discovery Rate During Real-Time Library Search Acquisitions

William D. Barshop[1], Chris C. McGann[2], Devin K. Schweppe[2], Jesse D. Canterbury[1]
[1]: Thermo Fisher Scientific, San Jose, CA, [2]: University of Washington, Seattle, WA

ABSTRACT

Estimation of False Discovery Rates for peptide-spectrum-matches (PSMs) made during acquisition from Real-Time Database Searching (RTS) serves to improve the number of quantified, confidently identified peptides in post-acquisition analysis. Here, we extend this strategy to examine Spectrum-Spectrum Matches (SSMs) for peptide identifications generated by Real-Time Library Search (RTLS). Classically, both RTS and RTLS depend on static user defined thresholds to determine whether a PSM or SSM is of sufficient quality to warrant triggering downstream scan behaviors. Here, we explore the extension of the False Detection Rate (FDR) model from RTS to the SSMs generated by RTLS. We examine useful score features generated by RTLS, and examine the utility of allowing these FDR estimates to guide triggering of additional scan behaviors.

MATERIALS AND METHODS

LC-MS data was acquired on a Thermo Scientific™ Easy nLC™-1200 coupled to a modified Thermo Scientific™ Orbitrap™ Tribrid™ mass spectrometer. Injections of the commercially available TKO yeast standard were separated on a Thermo Fisher Scientific™ EASY-Spray™ column. Spectral Libraries were generated from in silico predictions by INFERYS™ in Thermo Scientific™ Proteome Discoverer™ 3.0 and included decoy peptides to serve as negatives for training of the linear discriminant analysis (LDA) model[1]. The spectral library outputs were converted to mzVault™ format (.db) and used for Real-Time Library Search. Target peptide Spectrum-Spectrum Matches identified were pre-classified for use in training the LDA based on heuristic thresholds, as described previously[2-4]. Decoy SSMs and targets which did not meet these criteria were used as negatives to train the LDA throughout acquisition. For each run, once a sufficient quantity of target and decoy SSMs were accumulated, the LDA was trained and feature weights saved. The score threshold was found at which the user specified FDR is satisfied, and each incoming SSM evaluated by that standard. For every 1000 SSMs beyond this point, new weights were calculated via retraining of the LDA.

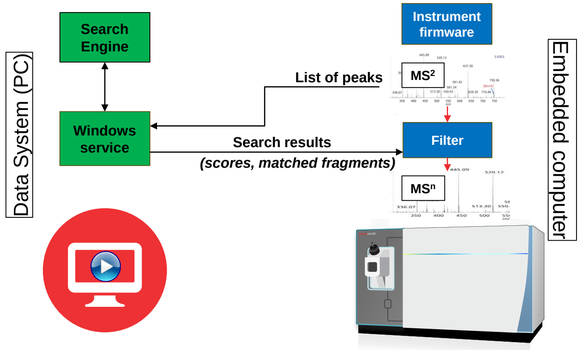

Figure 1. The Real-Time Search service receives each MS2 scan generated during acquisition. In turn, it produces and returns search results based on user provided filter parameters, and the design of the method. Upon receipt of the search results, the instrument will apply the relevant method logic to determine if the subsequent scan action will be executed and which peaks may be selected.

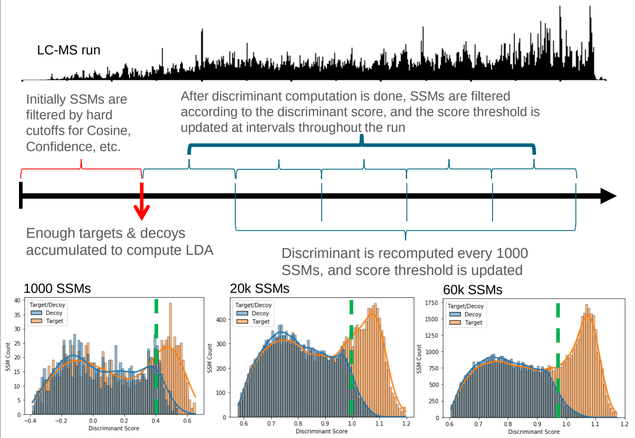

Figure 2. Evolution of target/decoy LDA score histograms throughout an acquisition. Green dashed lines show the approximate location of score thresholds calculated for 15% FDR.

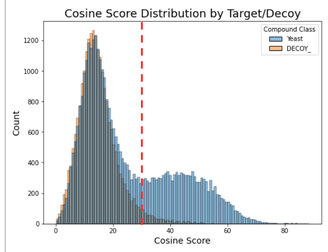

Figure 3. We heuristically defined the fixed threshold for Cosine score triggering based on a similar acquisition's target/decoy score distribution. The red dashed line denotes the utilized Cosine threshold of 30. This threshold was selected with the intention of rejecting the majority of decoy identifications while retaining the bulk of the target SSM distribution.

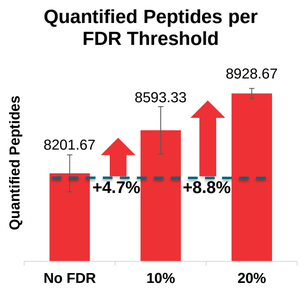

Figure 4. Acquired data was searched in Proteome Discoverer 3.0 with MSPepSearch against the same spectral library used for RTLS acquisition. Enablement of the FDR mode for RTLS led to an increase in quantified peptides. When using a FDR threshold of 20%, we increased the number of quantified peptides by approximate 8.8% over the no-FDR control runs. Data shown is an average of three acquisitions, with error bars showing +/- the standard deviation.

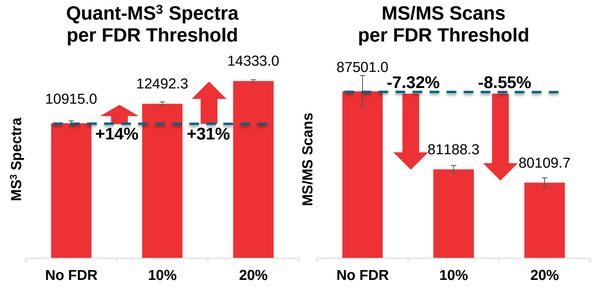

Figure 5. As the FDR threshold is increased to more permissive values, we observe an increase in the number of quantification SPS-MS³ scans acquired (left). We observe a concomitant decrease in the number of acquired MS² scans generated during these runs (right).

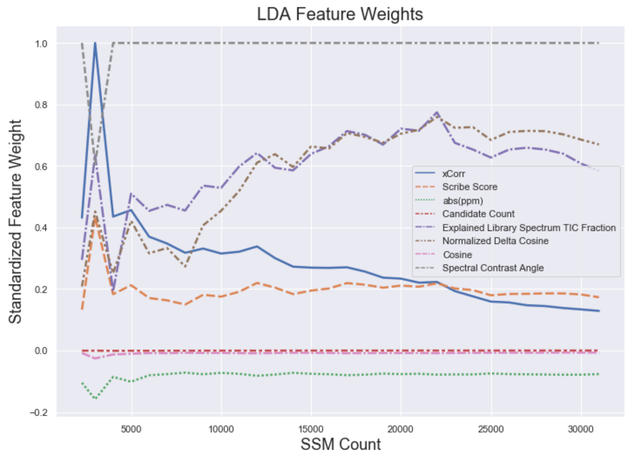

Figure 6. Standardized LDA feature weight evolution throughout an acquisition. Primarily, Spectral Contrast Angle contributes most heavily to the model, while Cosine and Candidate Count contribute the least. Interestingly, the Normalized Delta Cosine carries a relatively high weight throughout the run. Here, we also include the "Scribe Score", calculated as the log transformed inverse sum squared error of matched fragment intensities between the query and library spectra[4]. Early in the run, we observe some instability in the model weights which stabilize as more SSMs accumulate.

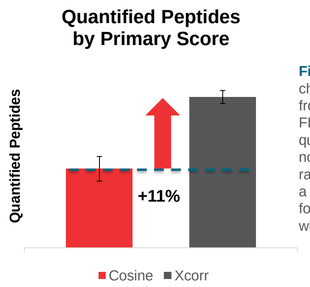

Figure 7. We investigated the impact of changing the primary ranking score for RTLS from Cosine to Xcorr in the absence of online FDR control. We found that the number of quantified peptides increased by 11% over the no FDR Cosine primary acquisition when ranking SSMs by Xcorr. This finding suggests a benefit to using Xcorr as the primary score for these linear trap MS² acquisitions which we will continue to investigate.

CONCLUSIONS

We find an increase in the number of triggered quantitative scans when "FDR-mode" is enabled, dependent upon the stringency of the manually defined score thresholds provided by the user. As with our FDR-mode implementation for RTS, this implementation allows only for additional triggering of scans when the calculated FDR thresholds would allow a SSM to pass, despite not reaching the user defined thresholds. In situations where a SSM would pass the user provided thresholds, but does not pass the FDR threshold, we allow the SSM to pass. In this way, FDR-mode acts only to rescue SSMs from inappropriate failure but not to safeguard against user thresholds which may be too lax. Together, the changes to RTLS in support of FDR mode provide added flexibility for bottom-up proteomics applications and help to ensure that peptides which can be confidently identified are also selected as triggers for dependent scans when an optimal score threshold is not obvious.

REFERENCES

1. Zolg, D. P., et al. (2021). "INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results." Rapid Commun Mass Spectrom: e9128.
2. Schweppe et al., J. Proteome Res. 2020, 19, 2026-2034; Erickson et al., J. Proteome Res. 2019, 18, 1299-1306.
3. Canterbury et al., Poster MP112, ASMS 2020.
4. Accord.NET Framework v. 3.8.0, available at http://accord-framework.net.
5. Searle, B. C., et al. (2023). "Scribe: Next Generation Library Searching for DDA Experiments." J Proteome Res 22(2): 482-490.

TRADEMARKS/LICENSING

© 2023 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.

For Research Use Only. Not for use in diagnostic procedures.

PO2023-116EN

ThermoFisher SCIENTIFIC