Machine learning-enabled Mascot Server API improves LFQ in client software like Thermo Proteome Discoverer

Ville Koskinen, Patrick Emery (Matrix Science Ltd, London, UK)*
** villek@matrixscience.com https://www.matrixscience.com/contact.html*

1. Background

Mascot Server identifies proteins from LC-MS/MS data. It is often used via software like Thermo Proteome Discoverer using a client API (application programming interface).

Mascot Server 3.1 can refine identification results using MS²PIP (spectrum prediction) and DeepLC (RT prediction). However, current software packages may not have any way to enable the functionality.

We have modified Mascot so new machine learning (ML) parameters can be encoded in the instrument definition. The enhanced client API encodes the posterior error probability (PEP) of a peptide match as -10*log10(PEP). The data format is 100% backwards compatible.

2. Questions

The client software may make (wrong) assumptions about the new peptide score.

- What settings require changing?
- Does refining with ML impact LFQ accuracy or precision?
- Does it provide higher protein coverage?

3. Methods

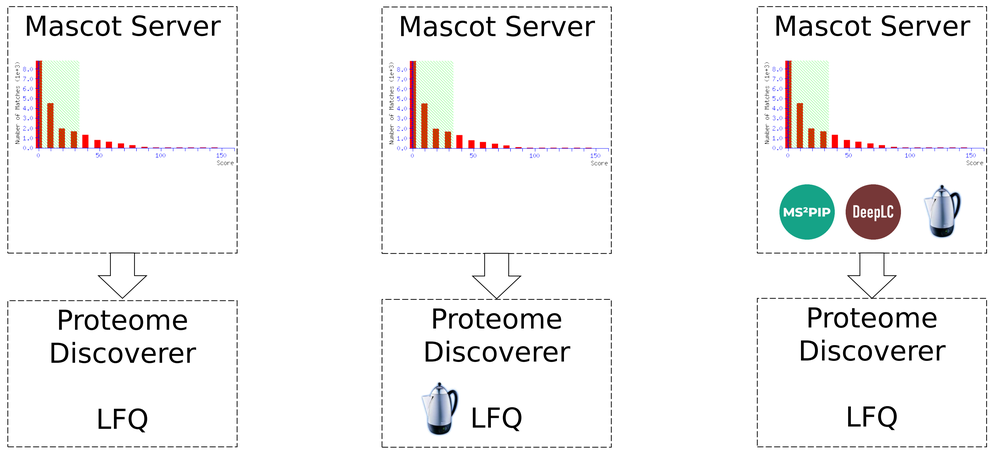
We reanalysed Thermo Orbitrap QE HF-X DDA raw files from PRIDE project PXD028735 (Puyvelde et al., Scientific Data 9:126, 2022), using Proteome Discoverer 3.2 and Mascot Server 3.1.

Samples A and B were run in triplicate. Human, yeast and *E. coli* proteins have expected log2 protein ratios 0, 1 and -2.

Accuracy: median log2 protein ratio per species.
Precision: median absolute deviation (MAD) from the median log2 ratio.

4. Results and conclusions

	No ML	Percolator node	Mascot ML
Instrument	ESI-TRAP	ESI-TRAP	HCD2019:hela_lumos_2h_psms



	Expected	Median	Precision	n	Median	Precision	n	Median	Precision	n
H. sapiens	0	0.009	0.069	2971	0.026	0.079	4260	0.030	0.077	4337
Yeast	1	1.09	0.09	1035	1.10	0.11	1733	1.10	0.11	1769
E. coli	-2	-1.87	0.15	235	-1.86	0.24	449	-1.84	0.24	460

Proteome Discoverer settings:
- Process each run separately (“By File”) if using DeepLC.
- Create multiconsensus report.
- Disable MudPIT scoring.

Full PD steps are on our website!

Enabling ML does not negatively impact LFQ accuracy. Minor change in precision.

Enabling MS²PIP and DeepLC provides deeper protein coverage.

MATRIX SCIENCE

COI: Ville Koskinen and Patrick Emery are directors and minority shareholders of Matrix Science Ltd.