

A unifying, spectrum-centric approach for the analysis of peptide tandem mass spectra



Daniel P. Zolg¹, Florian Seefried¹, Tobias Schmidt¹, Siegfried Gessulat¹, Michael Graber¹, Magnus Rathke-Kuhnert¹, Samia Ben Fredj¹, Patroklos Samaras¹, Bernhard Kuster², Mathias Wilhelm² & **Martin Frejno**¹

¹MSAID GmbH, Germany, ²Technical University of Munich, Germany

+ Introduction

1 CHIMERYYS: tackling chimeric peptide tandem mass spectra

"One peptide per spectrum" rarely exists, either by chance (DDA) or by design (DIA)

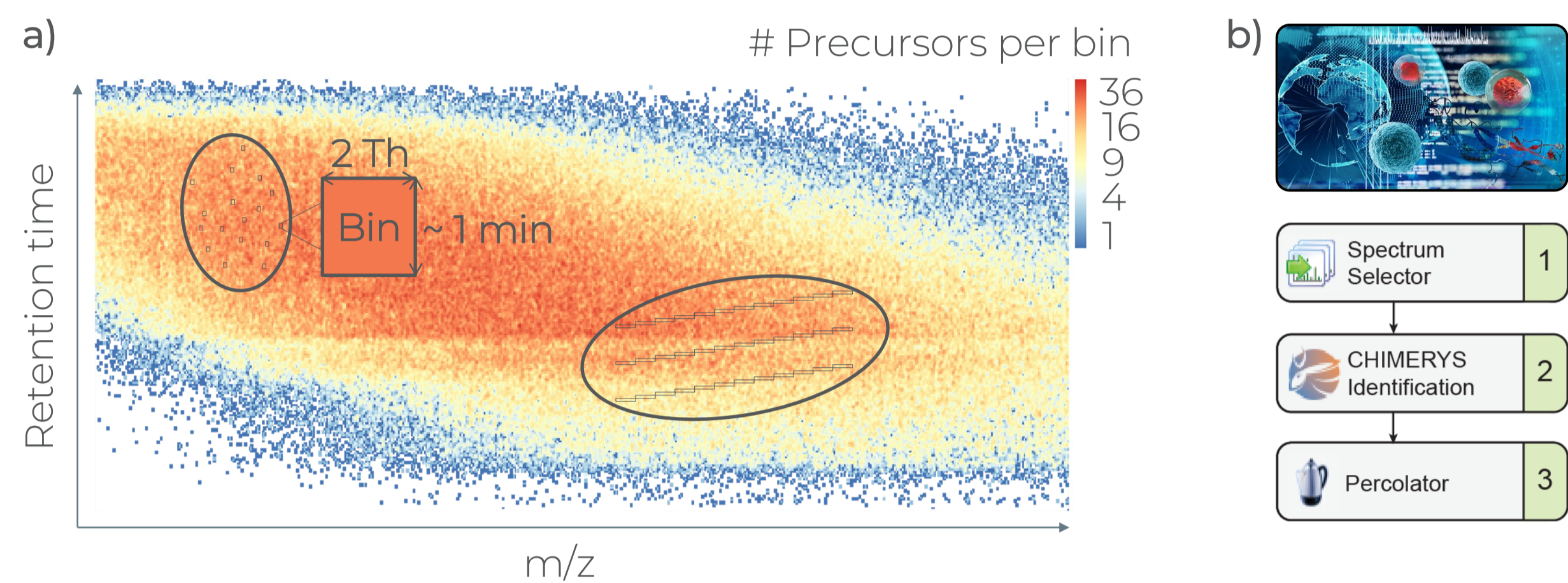


Figure 1 – a) Visualization of a human spectral library¹ b) CHIMERYYS workflow in PD 3.0

+ Results

2 Deconvolution of chimeric spectra doubles peptide identifications

CHIMERYYS IDs concur with various workflows and drastically increase PSMs

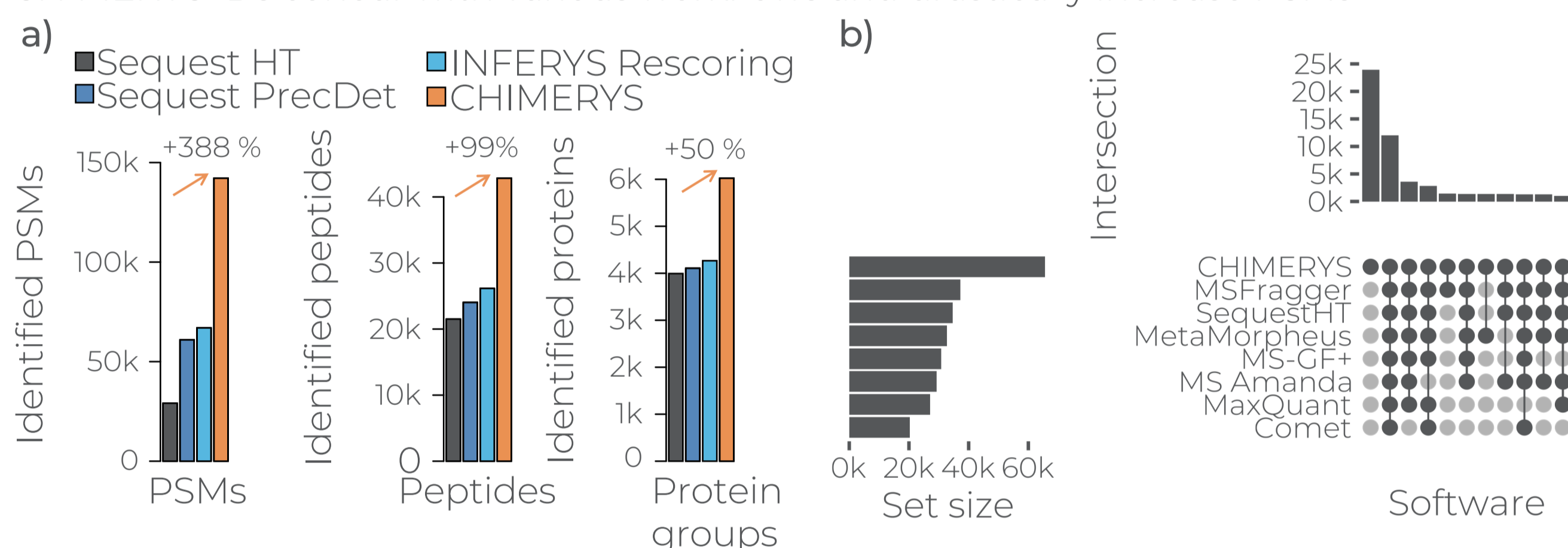


Figure 2 – a) CHIMERYYS vs Sequest variants b) CHIMERYYS vs other software

3 CHIMERYYS excels both for deep-proteomes and fast gradients

Fractionated samples & high-throughput applications profit from ML-aided scoring

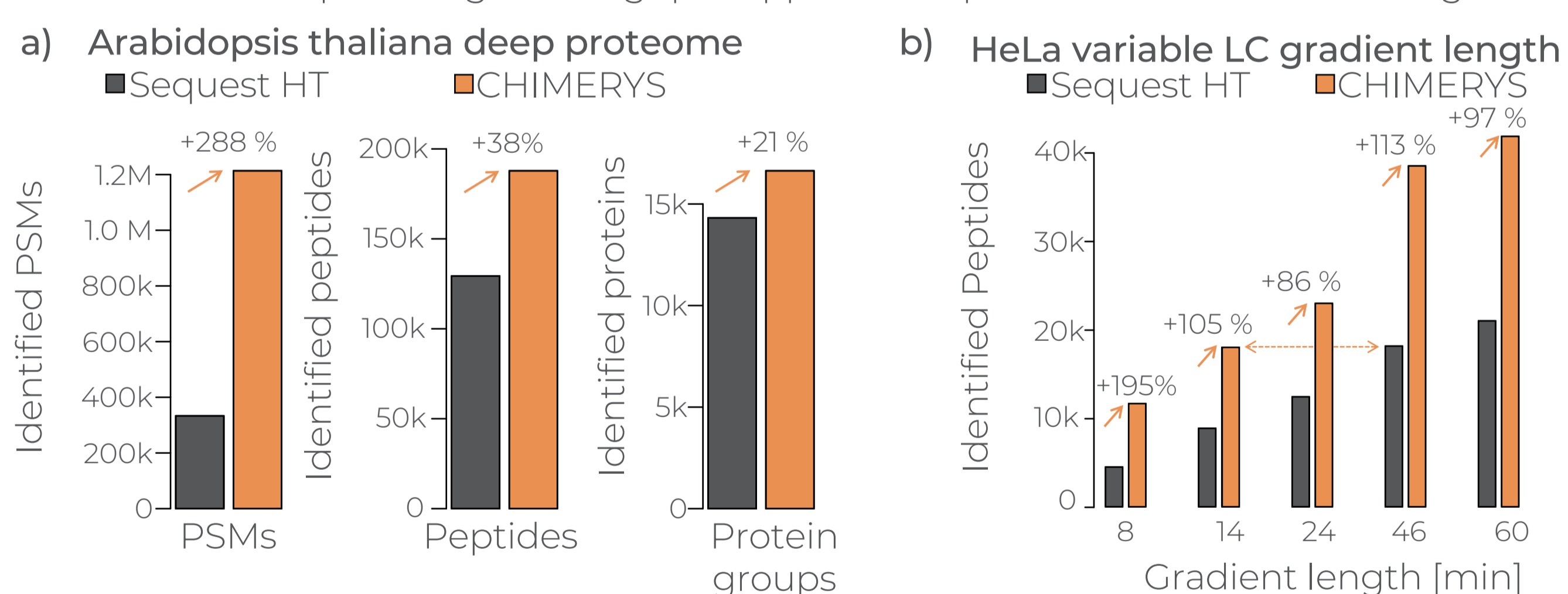


Figure 3 – a) Deeply fractionated dataset b) Improvement of IDs over gradient length

4 Entrapment experiments with wide-window DDA data

1x human & 9x human shuffled fasta: accurate FDR estimation in wide-window DDA

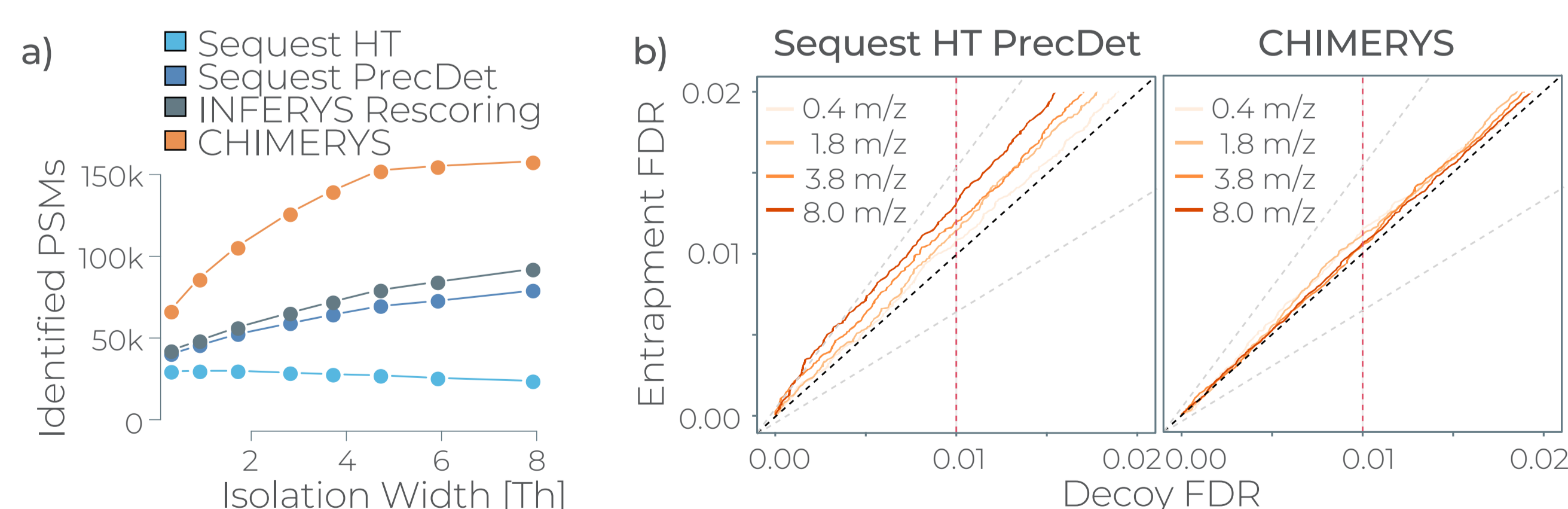


Figure 4 – a) Increasing isolation windows b) Entrapment on wide-window DDA data

5 Comparison to peptide-centric quantification on DIA data²

Different scoring approaches show correlation of quantification

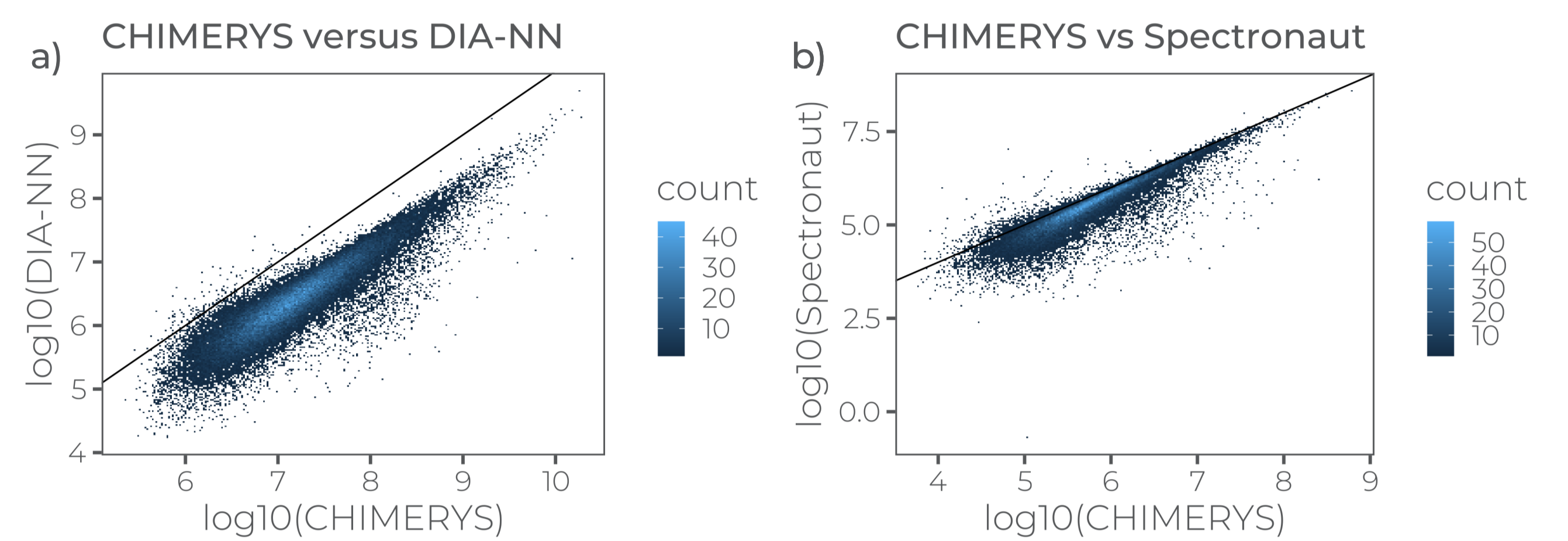


Figure 5 – a) Comparison to DIA-NN b) Comparison to Spectronaut 16

6 High overlap of IDs and accurate quantification of DIA data

Different scoring approaches show overlap of IDs & CHIMERYYS quantifies accurately

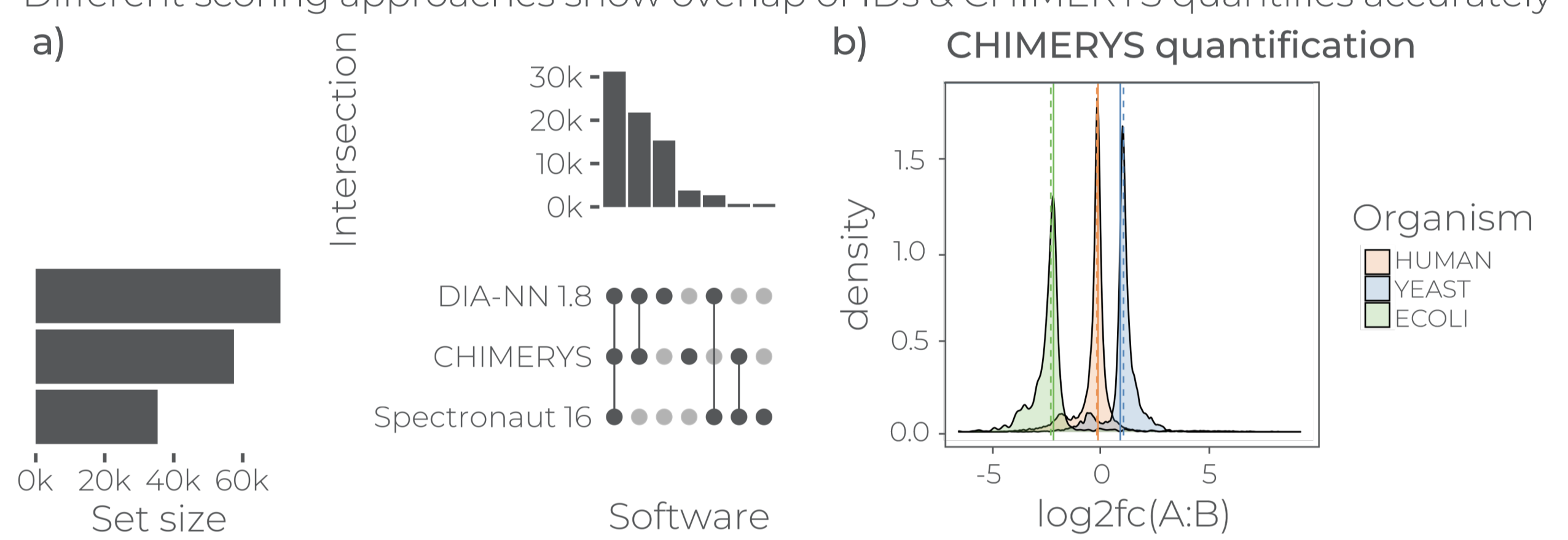


Figure 6 – a) CHIMERYYS vs other DIA software b) Recovery of known ratios

7 Aggregated, coefficient-based quantification matches Skyline

Automated scoring is correlated to expert-performed, manual quantification of XICs

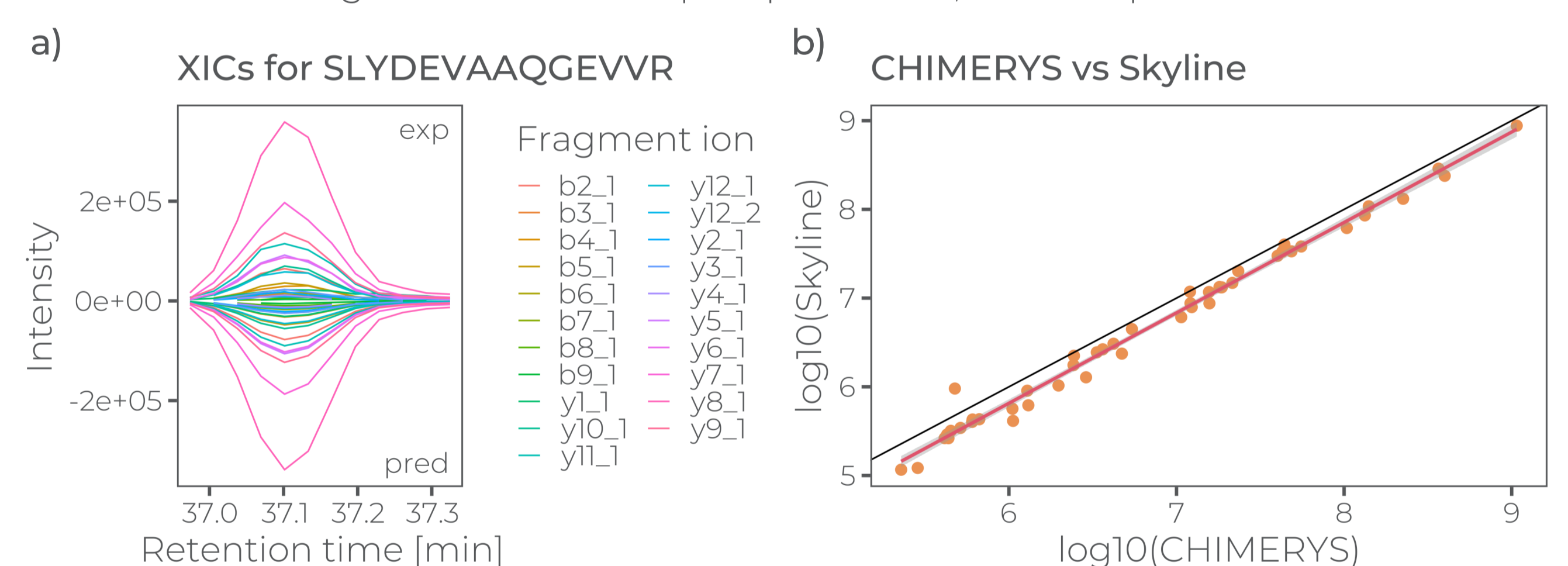


Figure 7 – a) Mirror XICs validate IDs b) High correlation over 3.5 orders of magnitude

+ Conclusions

- + Deep-learning based predictions enable systematic spectrum-centric data analysis
- + A novel deconvolution algorithm yields striking results for chimeric DDA spectra
- + Deconvolution concept can be translated from wide-window DDA to DIA and PRM
- + Coefficient-based quantification matches current MS2-based quantification
- + We predict a bright future for ML-based applications in proteomics

+ About MSAID

For more information please visit: www.msaid.de/conferences/ushupo2023

References

- ¹Zhu, T et. al, 2020. DPHL: A DIA Pan-human Protein Mass Spectrometry Library for Robust Biomarker Discovery
²Van Puyvelde, B. et al, 2022. A comprehensive LFQ benchmark dataset on modern day acquisition strategies in proteomics

CHIMERYYS®, INFEREYS® and MSAID® are registered trademarks of MSAID GmbH. Thermo Scientific™ Proteome Discoverer™ software is a trademark of Thermo Fisher Scientific Inc. Sequest™ is a trademark of the University of Washington