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

CHIMERYS
BY MSAID

Daniel P. Zolg[1], Florian Seefried[1], Tobias Schmidt[1], Siegfried Gessulat[1], Michael Graber[1], Magnus Rathke-Kuhnert[1], Samia Ben Fredj[1], Patroklos Samaras[1], Bernhard Kuster[2], Mathias Wilhelm[2] & Martin Frejno[1]
[1]MSAID GmbH, Germany, [2]Technical University of Munich, Germany

+ Introduction

1 CHIMERYS: tackling chimeric peptide tandem mass spectra

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

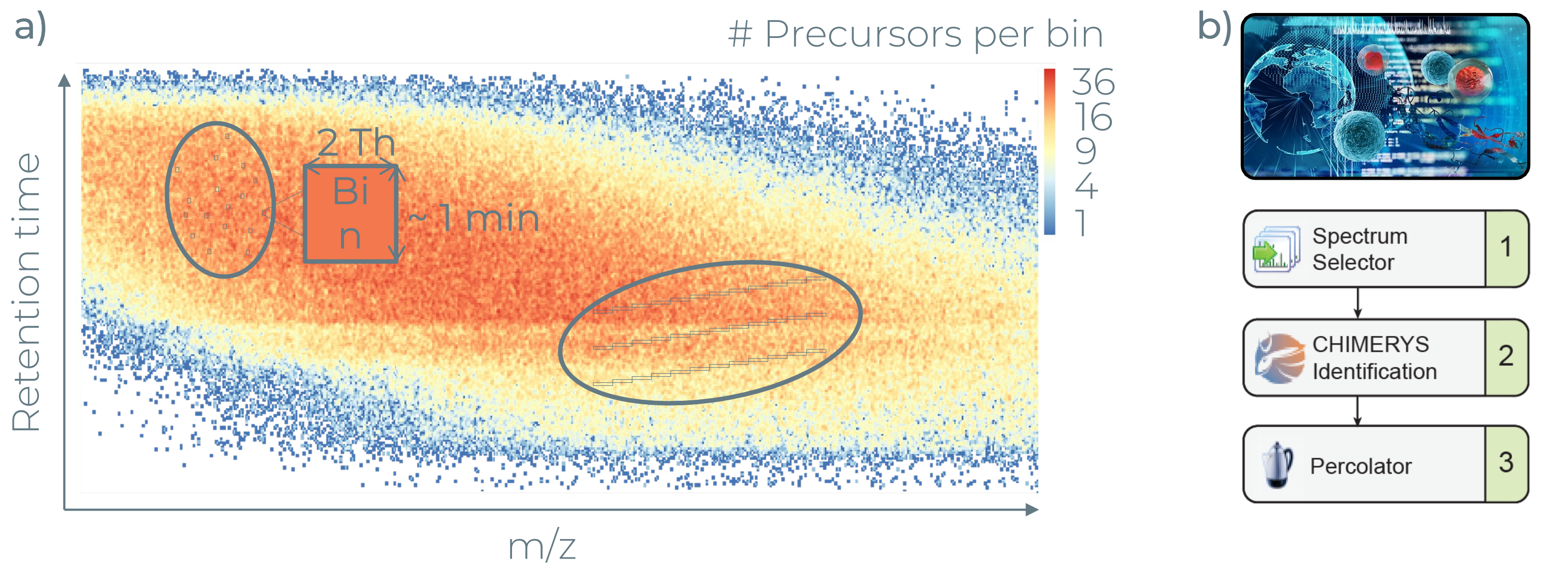

Figure 1 – a) Visualization of a human spectral library[1] **b)** CHIMERYS workflow in PD 3.0

+ Results

2 Deconvolution of chimeric spectra doubles peptide identifications

CHIMERYS IDs concur with various workflows and drastically increase PSMs

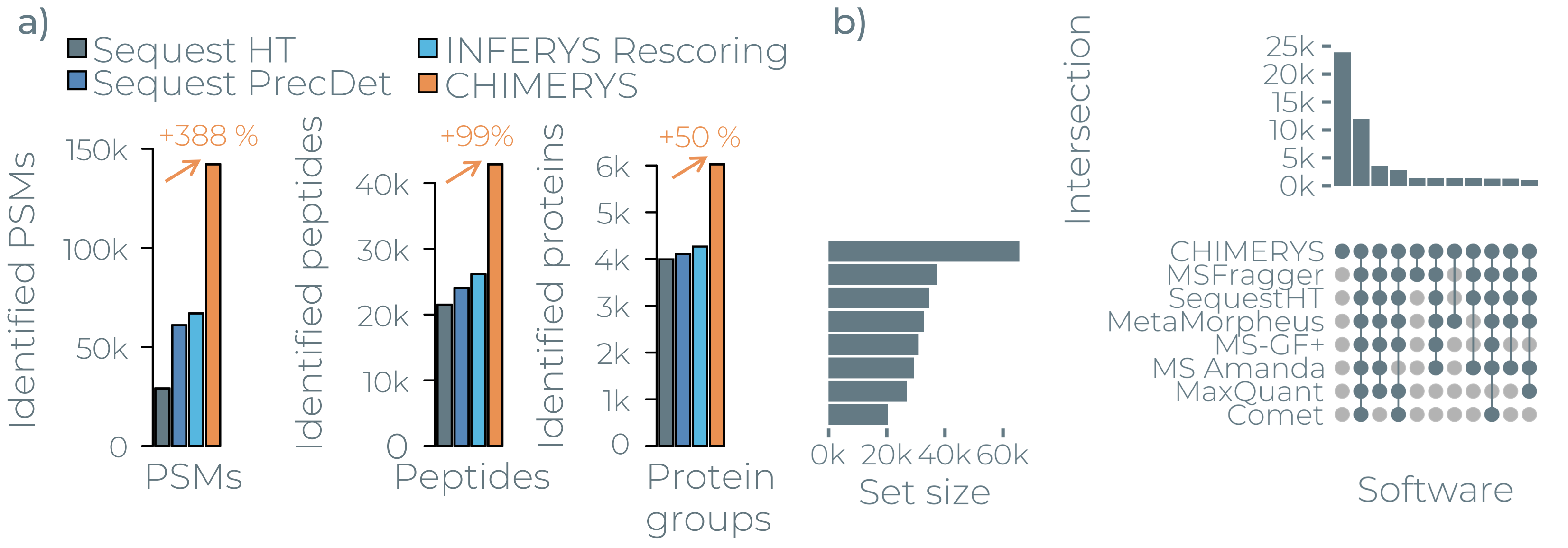

Figure 2 – a) CHIMERYS vs Sequest variants **b)** CHIMERYS vs other software

3 CHIMERYS excels both for deep-proteomes and fast gradients

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

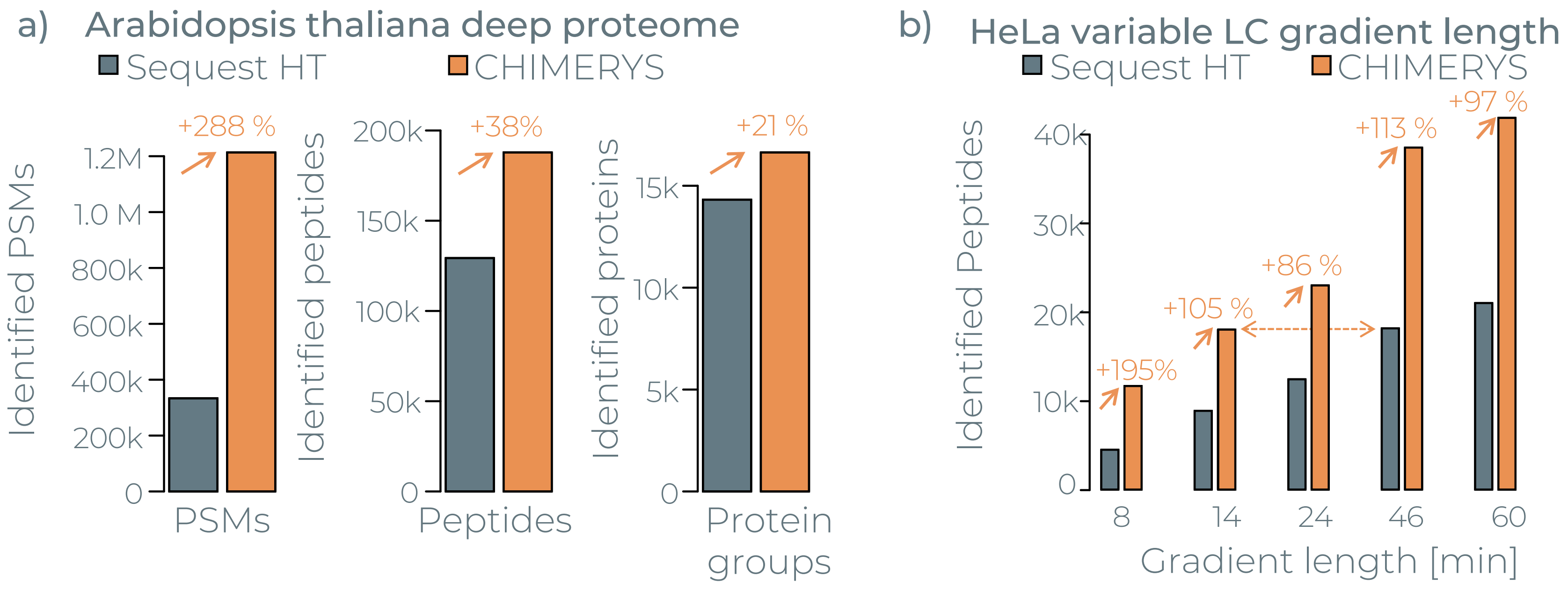

Figure 3 – a) Deeply fractionated dataset **b)** Improvement of IDs over gradient length

4 Entrapment experiments with wide-window DDA data

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

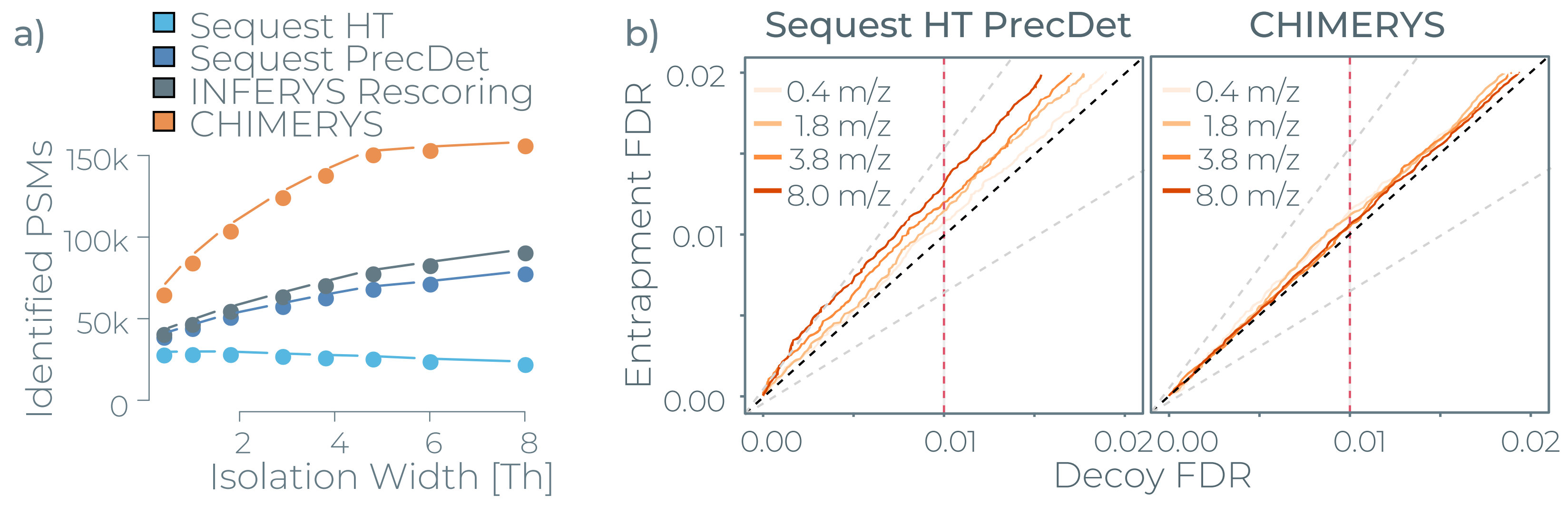

Figure 4 – a) Increasing isolation windows **b)** Entrapment on wide-window DDA data

5 Comparison to peptide-centric quantification on DIA data[2]

Different scoring approaches show correlation of quantification

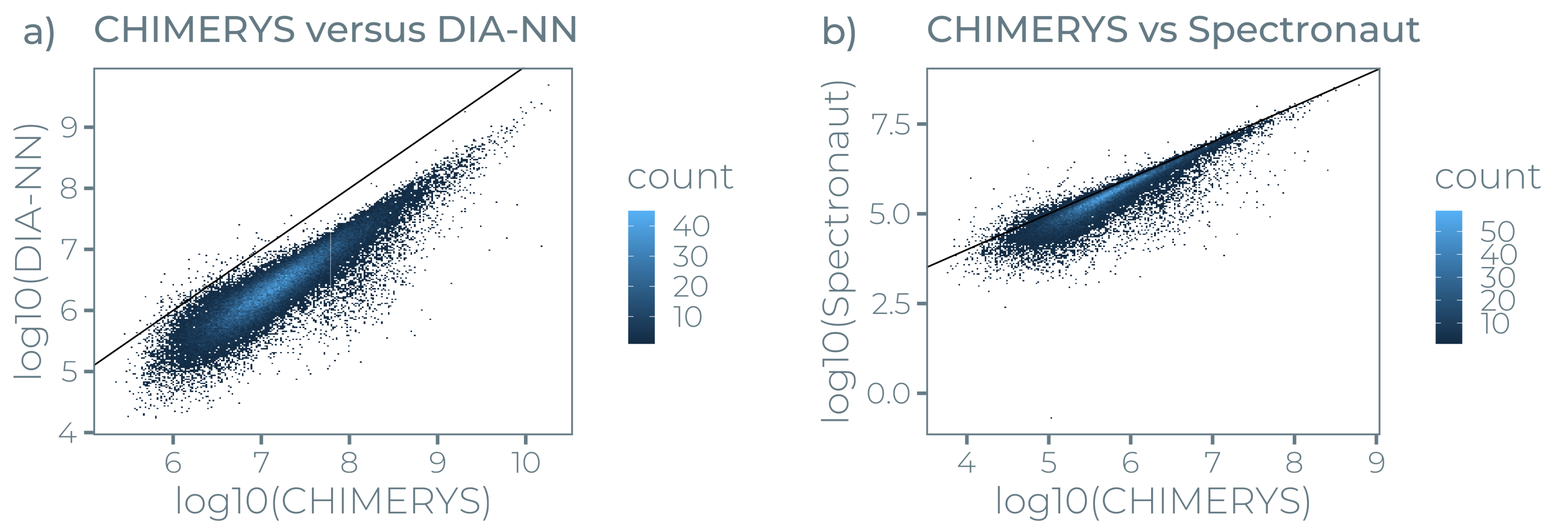

Figure 5 – a) Comparison to DIA-NN **b)** Comparison to Spectronaut 16

6 High overlap of IDs and accurate quantification of DIA data

Different scoring approaches show overlap of IDs & CHIMERYS quantifies accurately

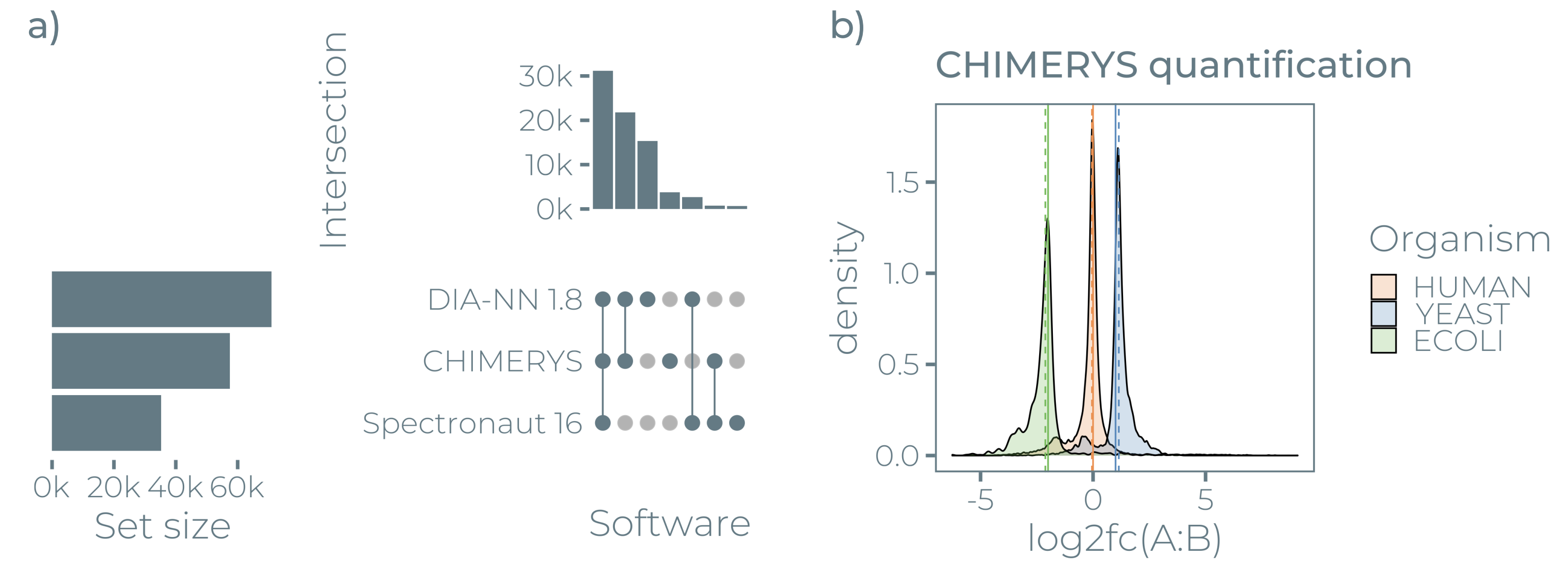

Figure 6 – a) CHIMERYS vs other DIA software **b)** Recovery of known ratios

7 Aggregated, coefficient-based quantification matches Skyline

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

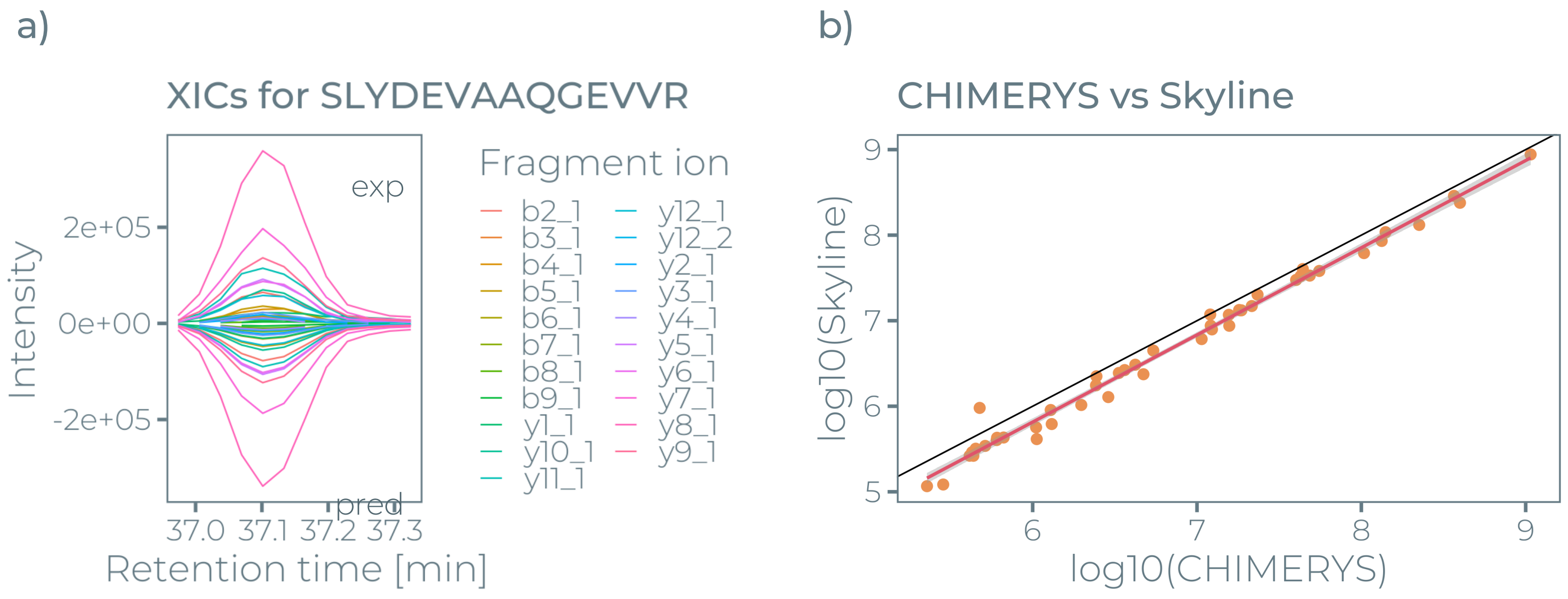

Figure 7 – a) Mirror XICs validate IDs **b)** High correlation over 3.5 orders of magnitude

+ 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

+ Related Content

Poster PP03.27 (Wednesday 7th of December | AM Session 11:45 AM - 1:00 PM): Digging deeper into phosphoproteomes through AI-driven deconvolution of chimeric spectra.

MSAID BOOTH #211 | https://www.msaid.de/conferences/hupo2022

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