



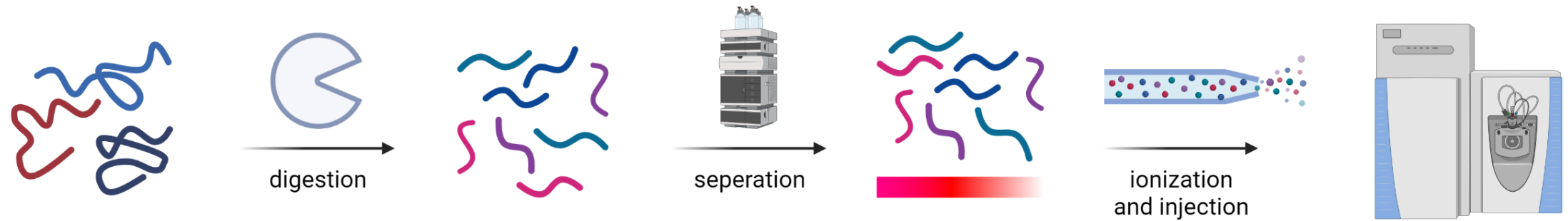
ENABLING NOVEL AND CHALLENGING PROTEOMICS WORKFLOWS WITH MS²RESCORE

Ralf Gabriels

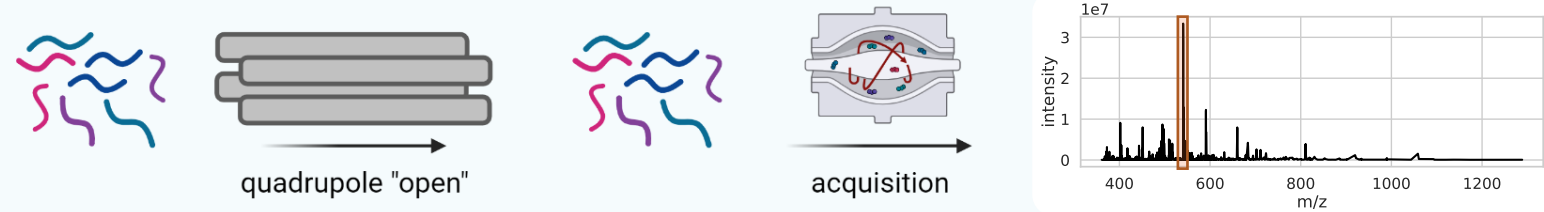
EuPA ECR Day 2022

28/06/2022

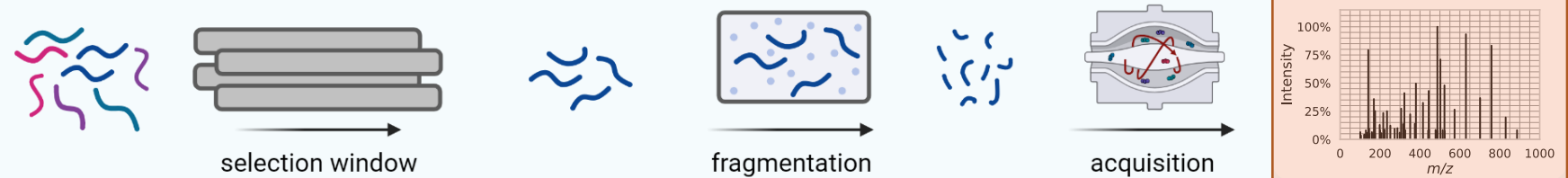
Mass spectrometry-based proteomics recap



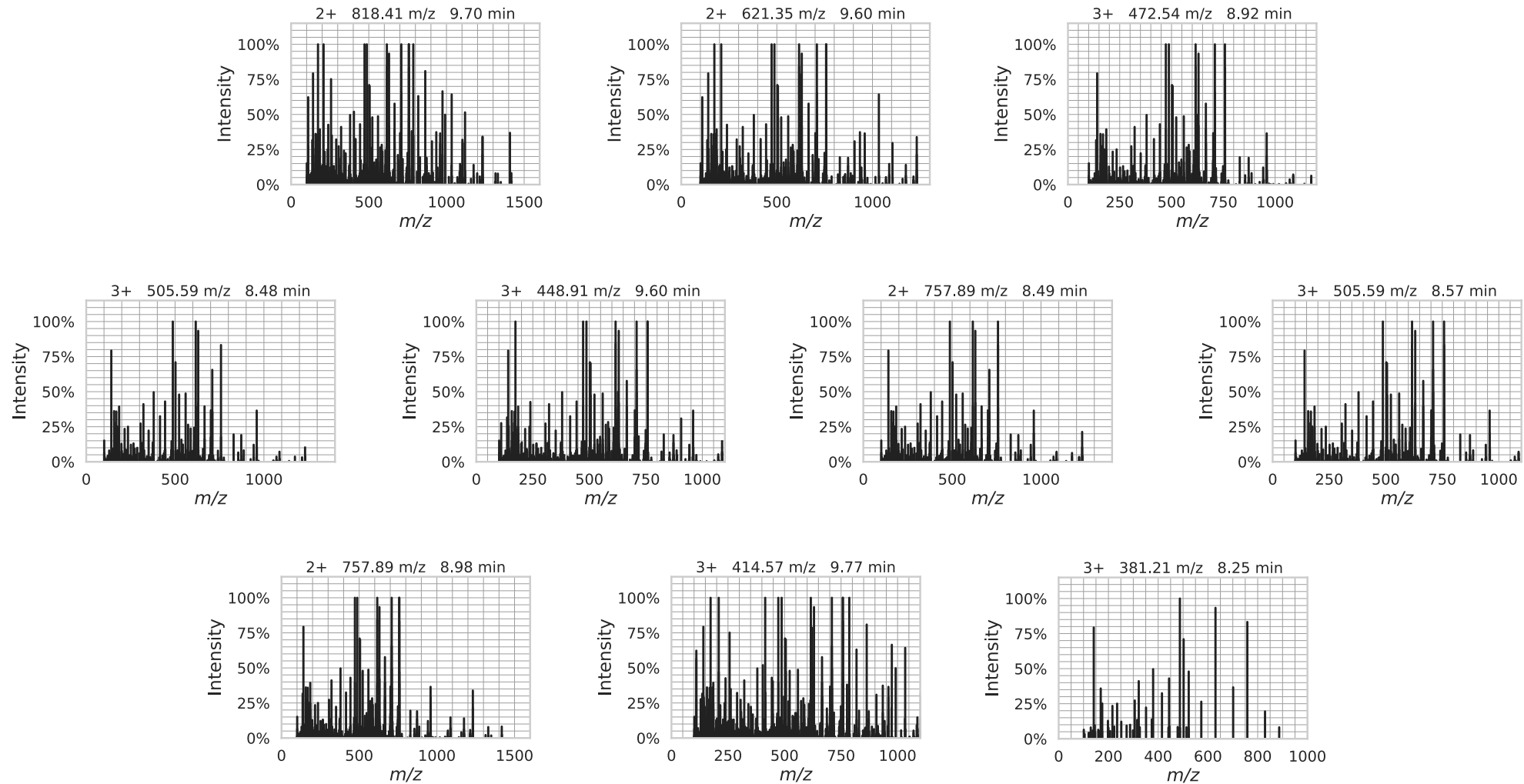
MS1 survey scan:
Find precursors (= full peptides)



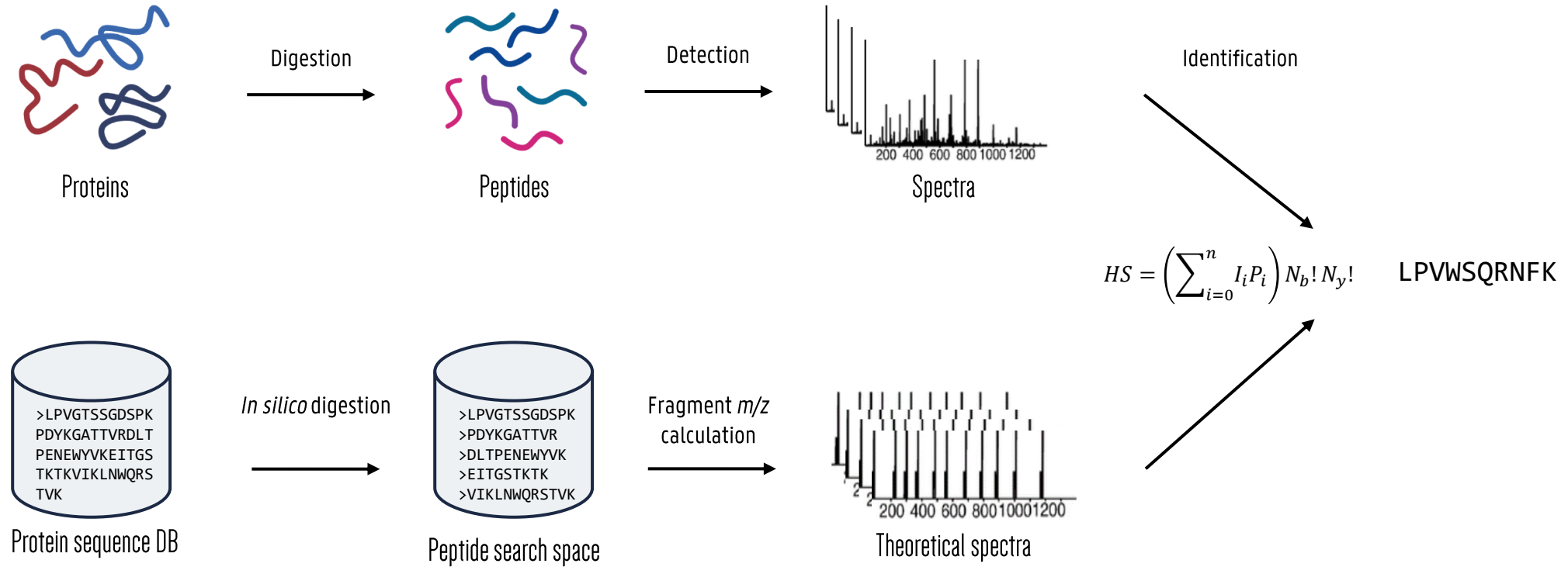
MS2 scan:
Get fragmentation spectrum for peptide at selected m/z window



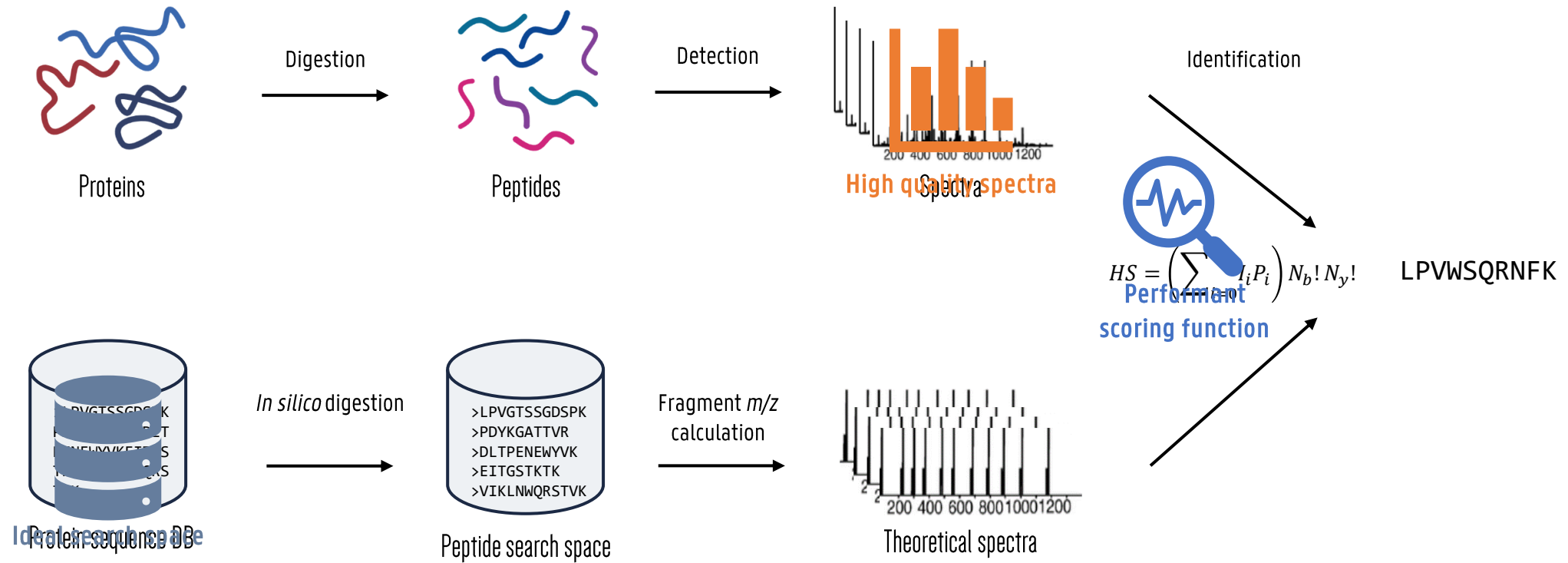
How do we link MS2 spectra back to the original proteins?



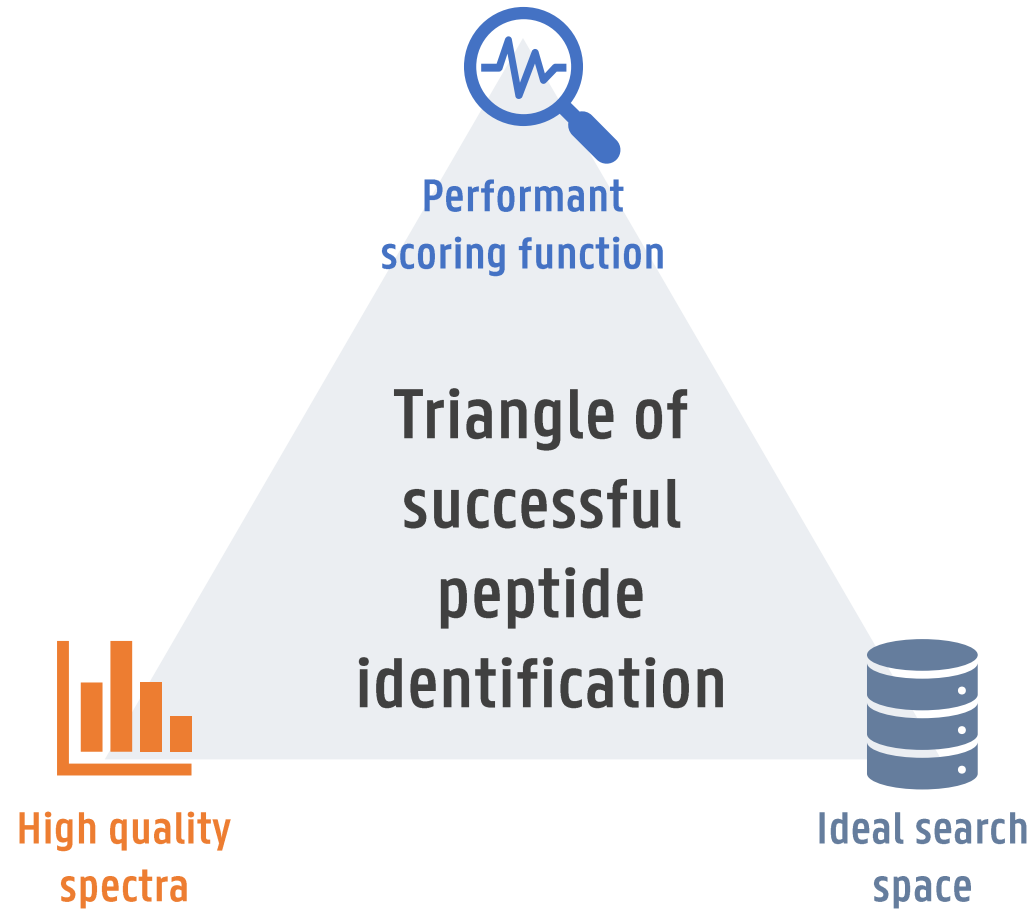
Proteomics search engines replicate the LC-MS steps *in silico*



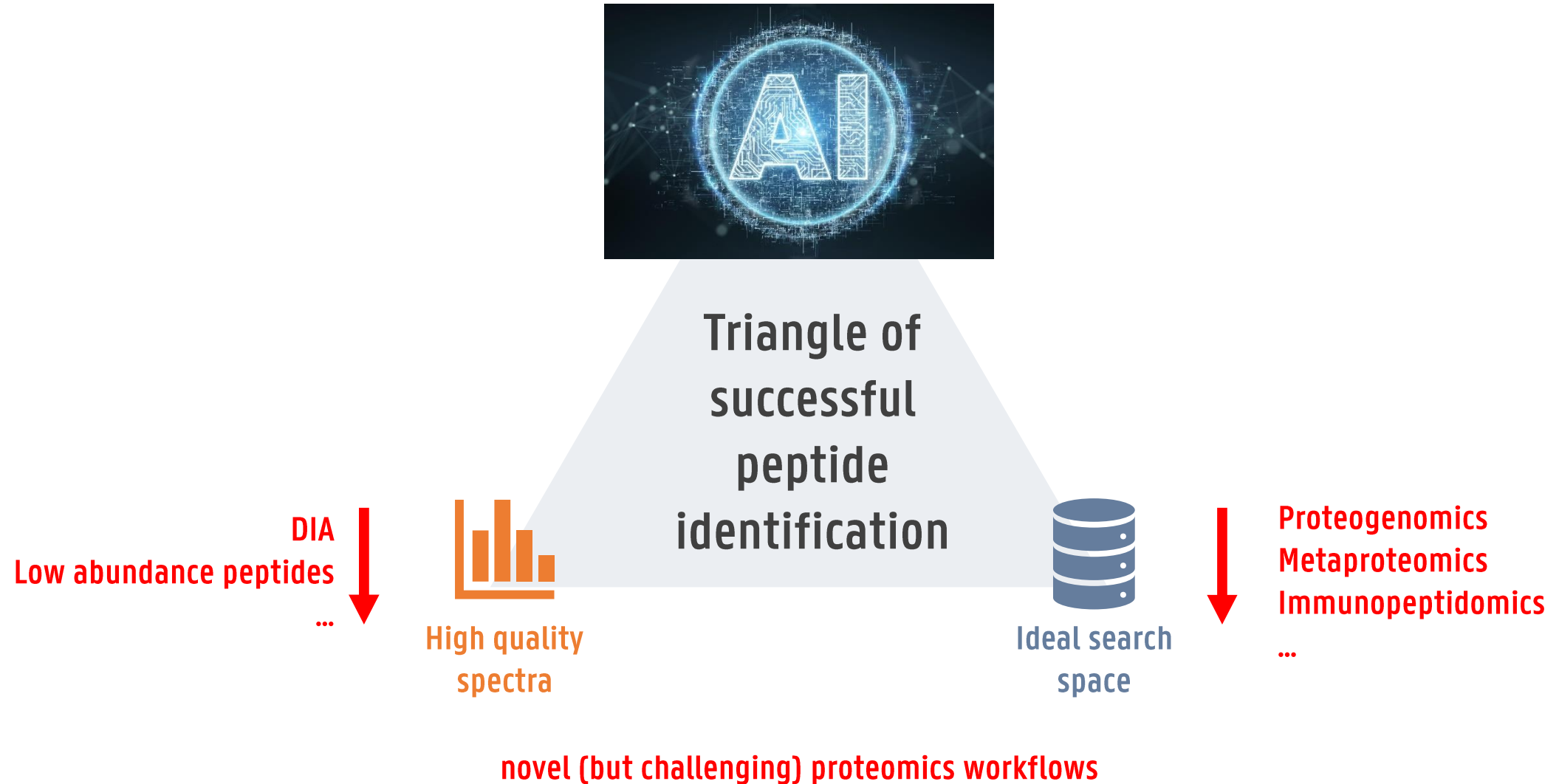
Key elements to successful peptide spectrum identification?



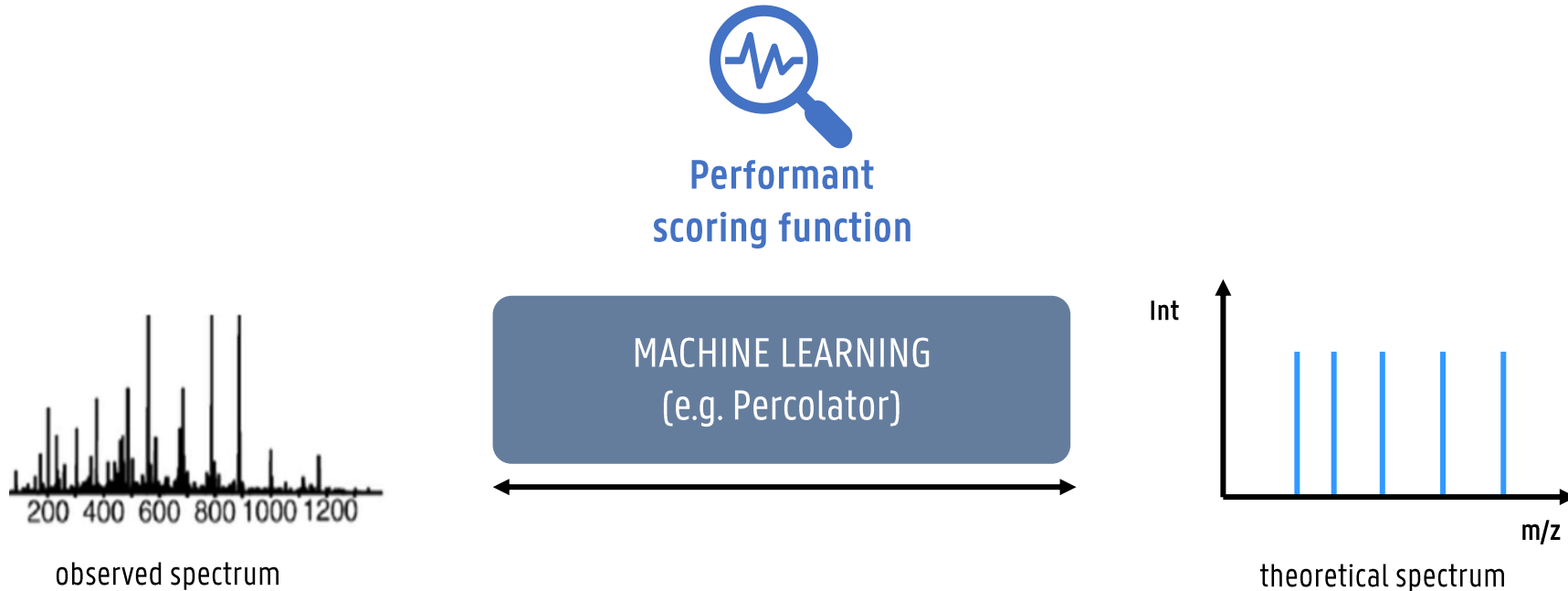
Key elements to successful peptide spectrum identification?



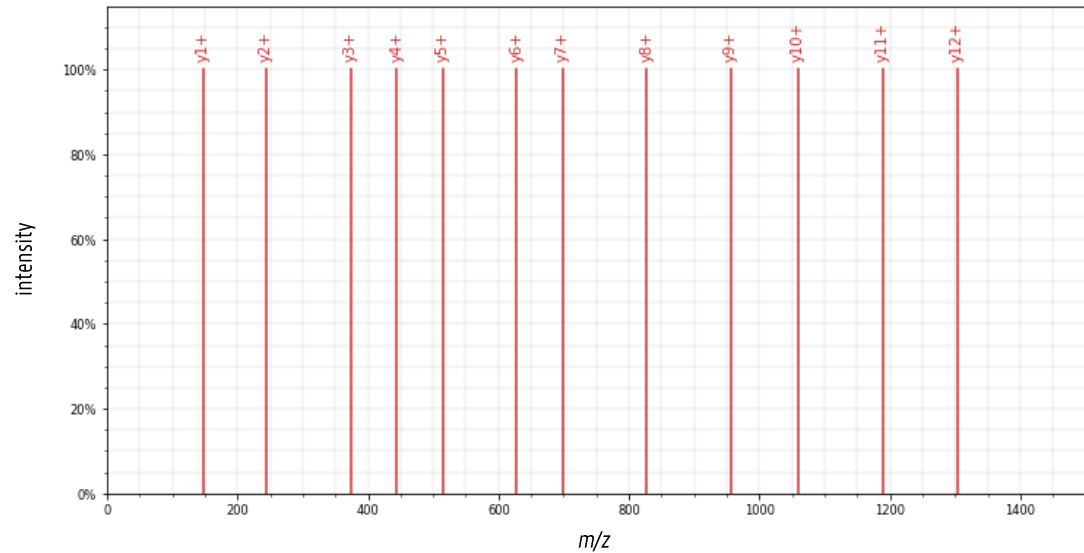
Key elements to successful peptide spectrum identification?



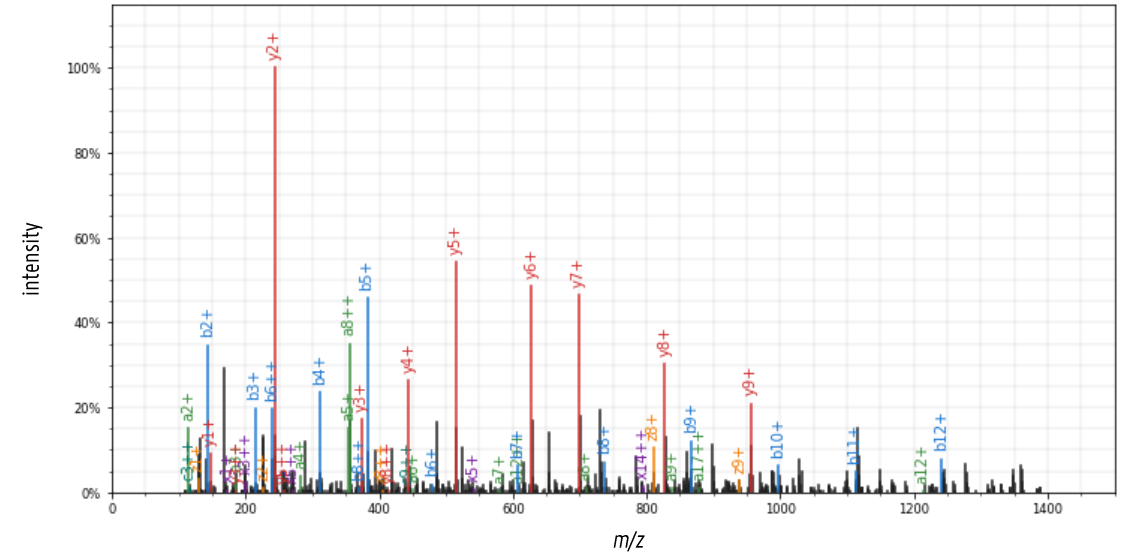
Machine learning can generate a dynamically optimized scoring function, tailored to each data set



MS²PIP brings a more realistic view on peptide fragmentation to PSM scoring

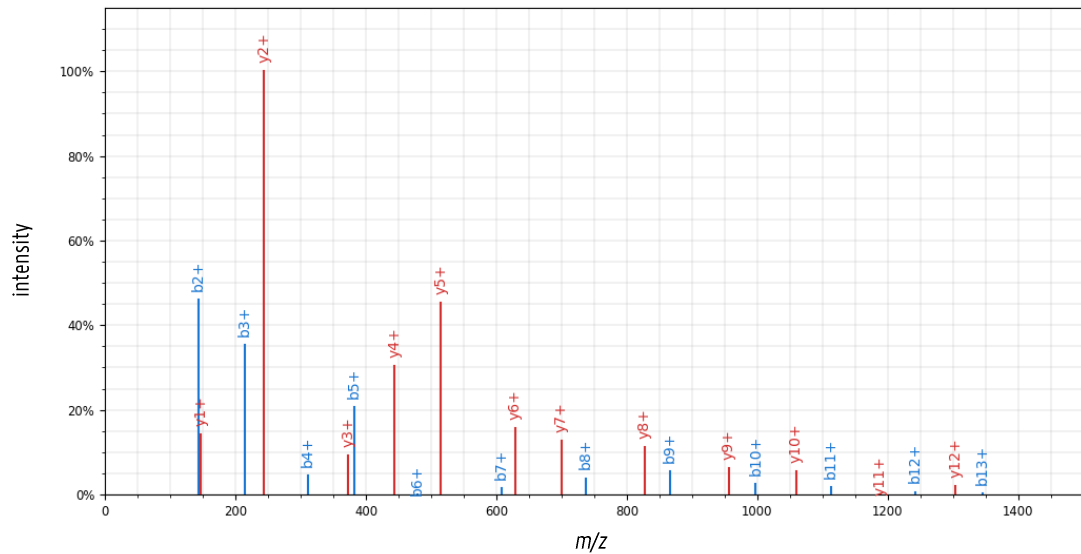
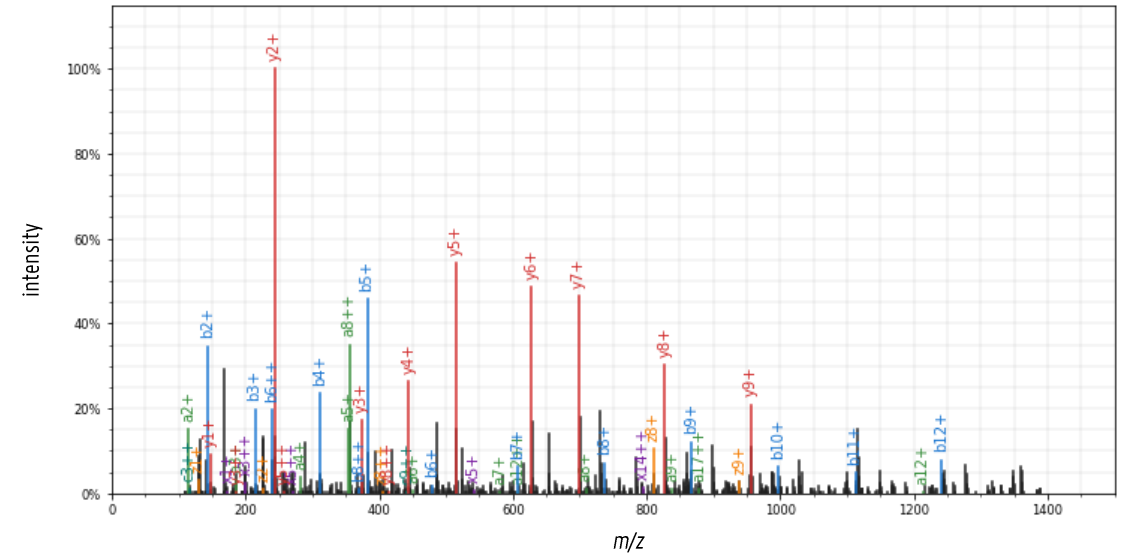


theoretical spectrum



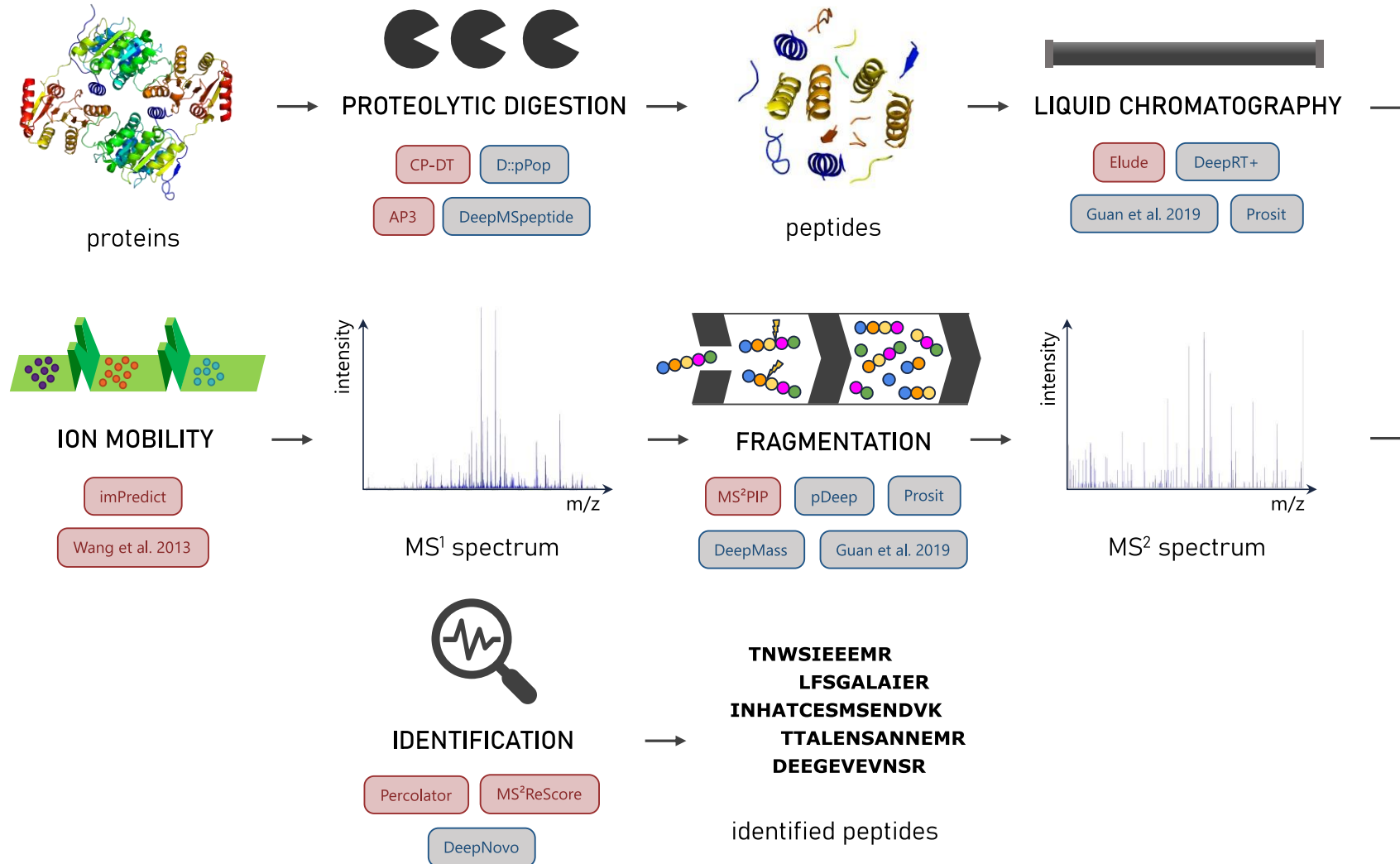
observed spectrum

MS²PIP brings a more realistic view on peptide fragmentation to PSM scoring

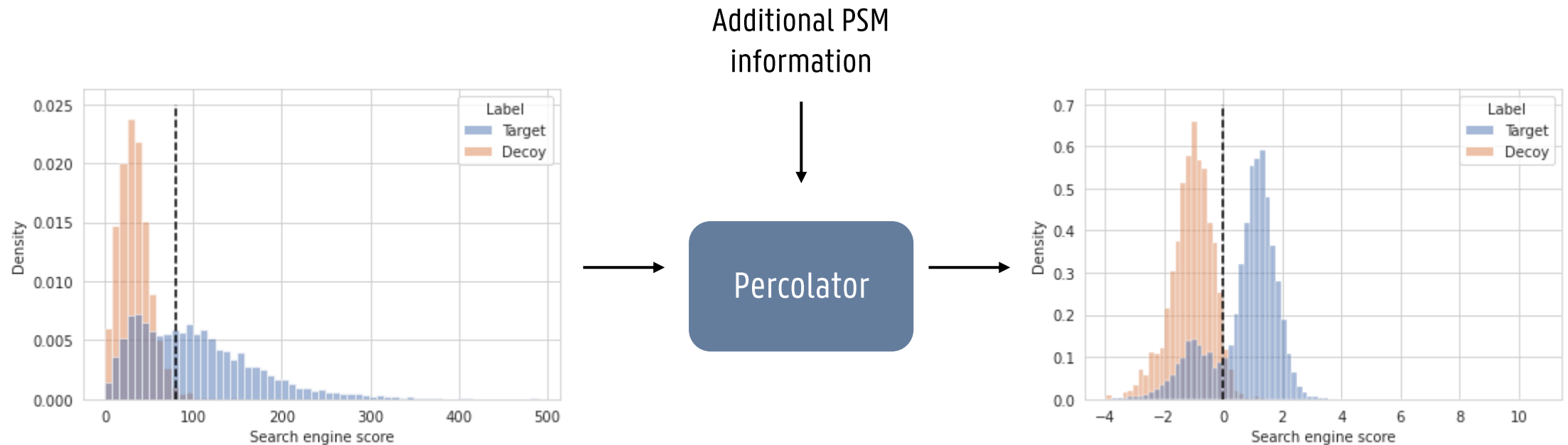
MS²PIP-predicted spectrum

observed spectrum

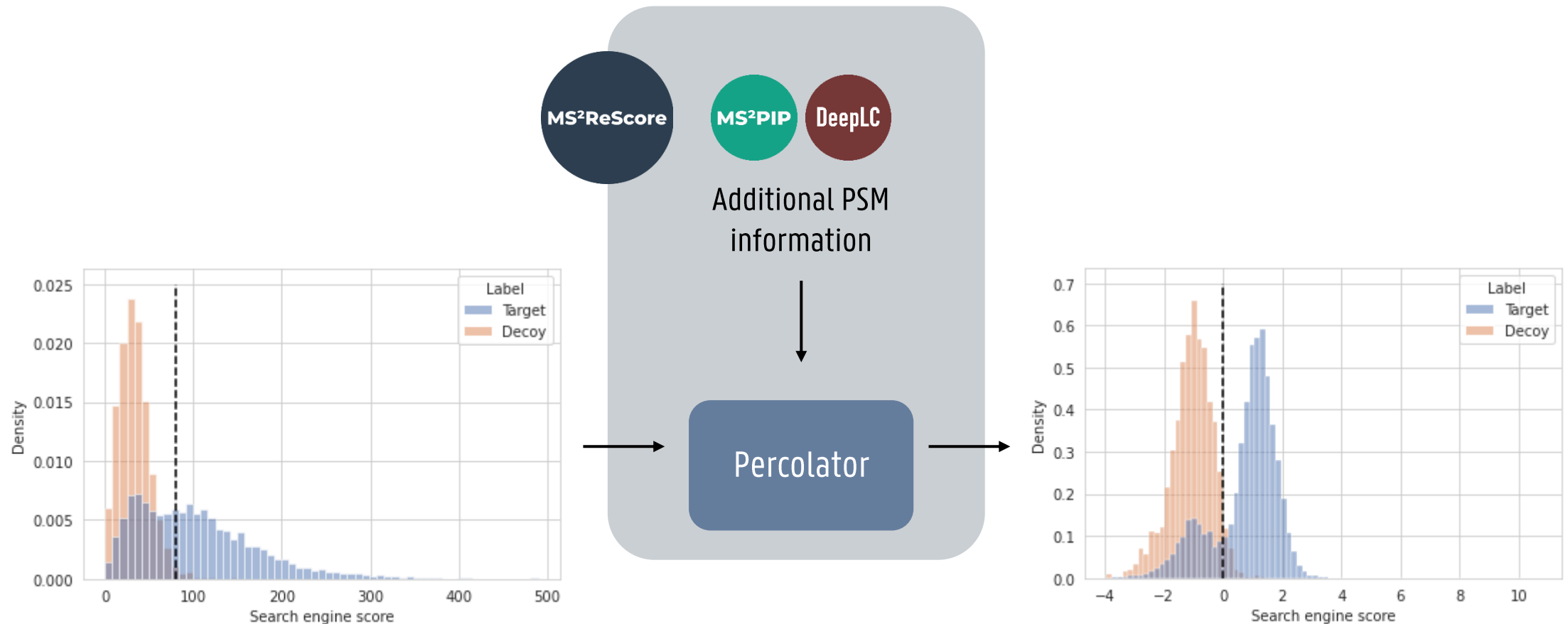
Nearly the whole LC-MS workflow can be modeled



Machine learning-based information can be passed to Percolator for improved rescoring



Machine learning-based information can be passed to Percolator for improved rescoring

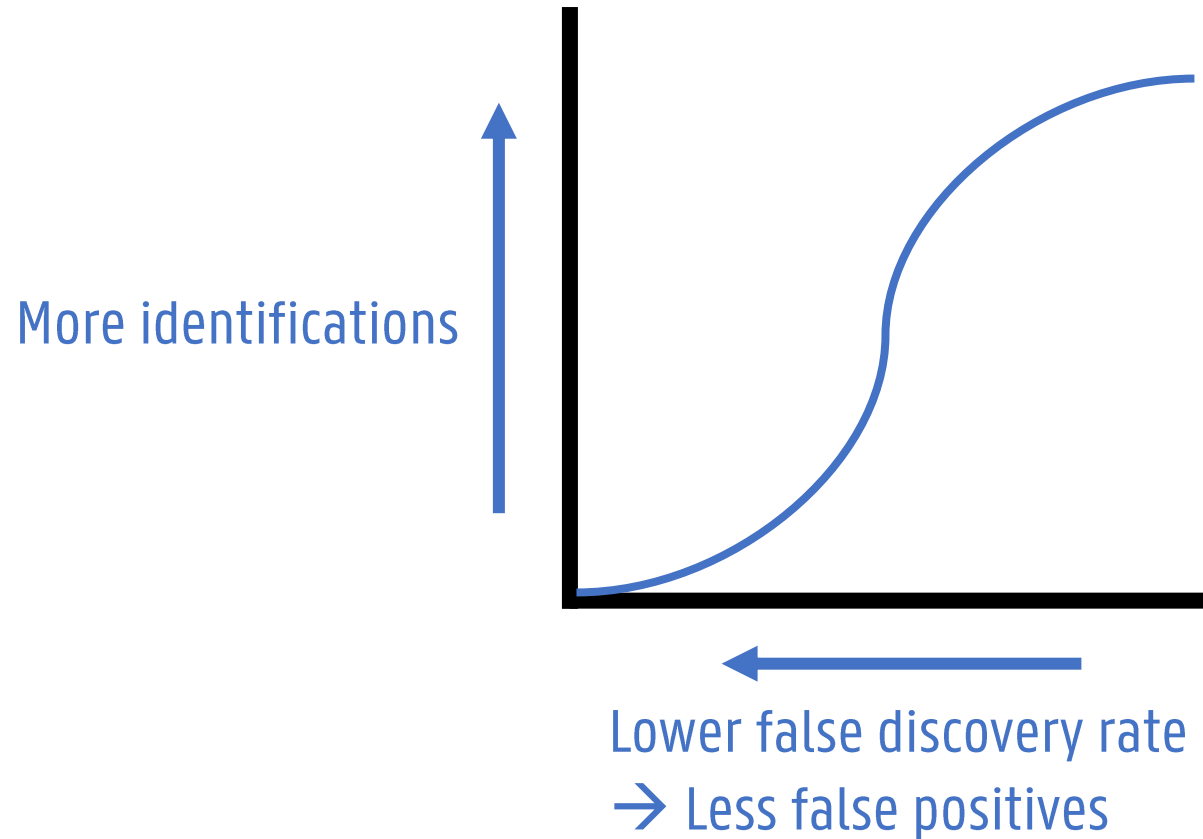


Rescoring with MS²PIP: Ana S C. Silva et al. (2019) Bioinformatics. doi:10.1093/bioinformatics/btz383

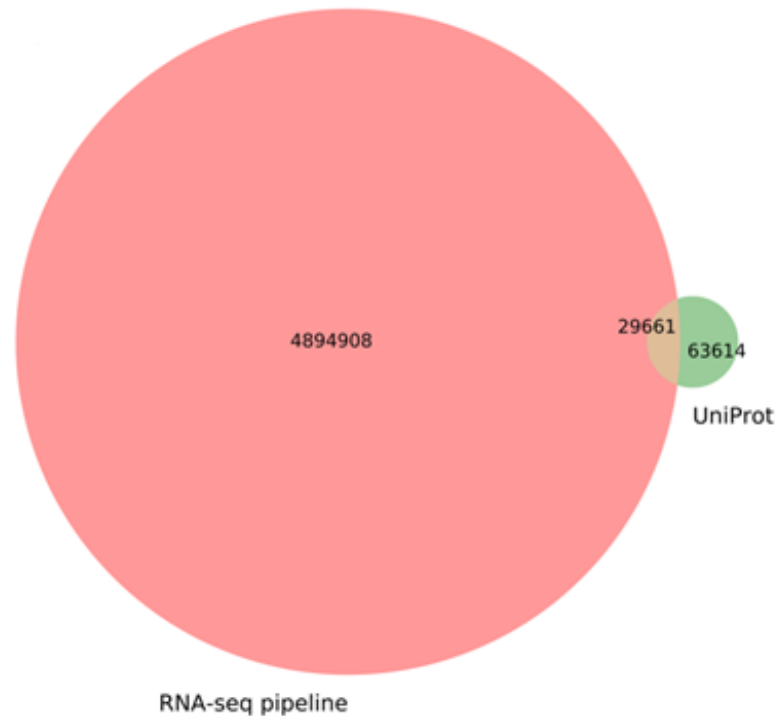
DeepLC: Robbin Bouwmeester et al. (2021) Nature Methods. doi:10.1038/s41592-021-01301-5

MS²Rescore: Arthur Declercq, [...] Ralf Gabriels (2021) bioRxiv. doi:10.1101/2021.11.02.466886

Intermezzo: Explaining FDR/identification rate plots



MS²Rescore in proteogenomics: Higher ID rate, at a 10-fold lower FDR threshold

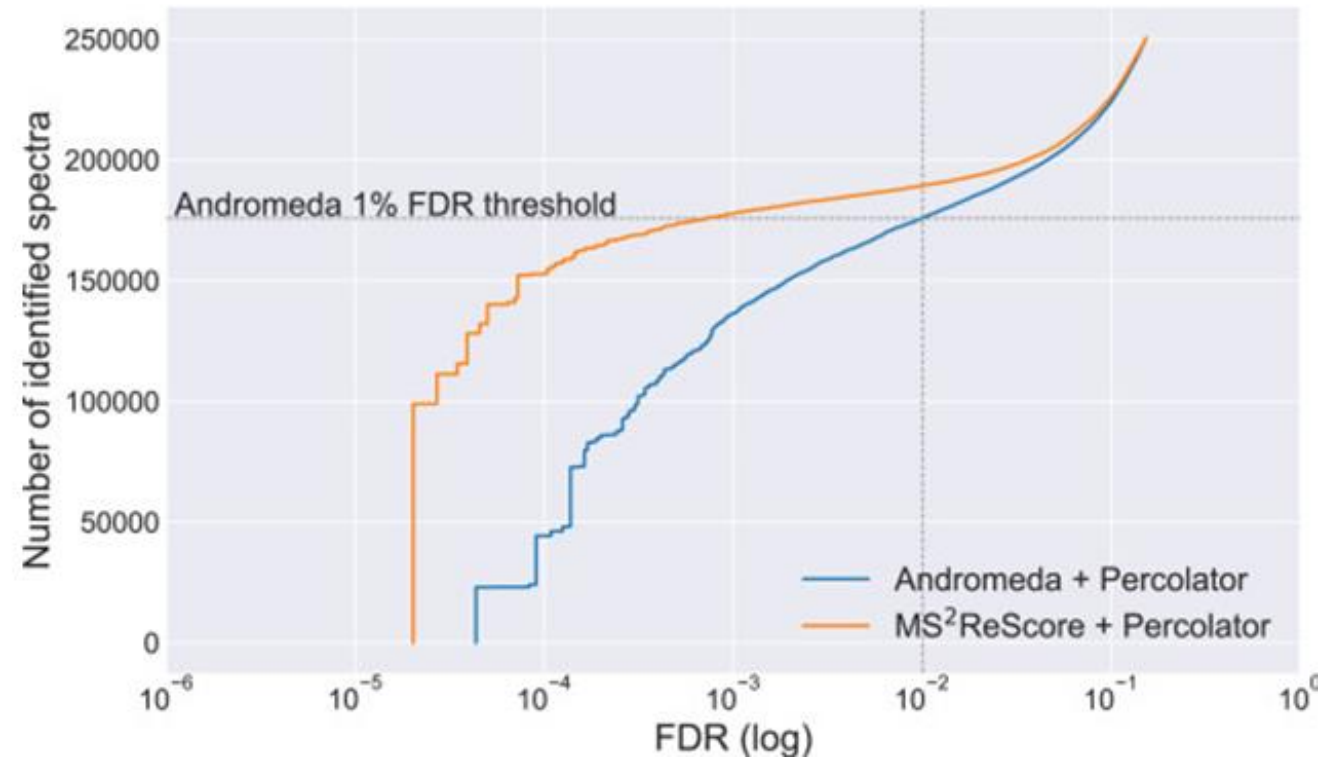
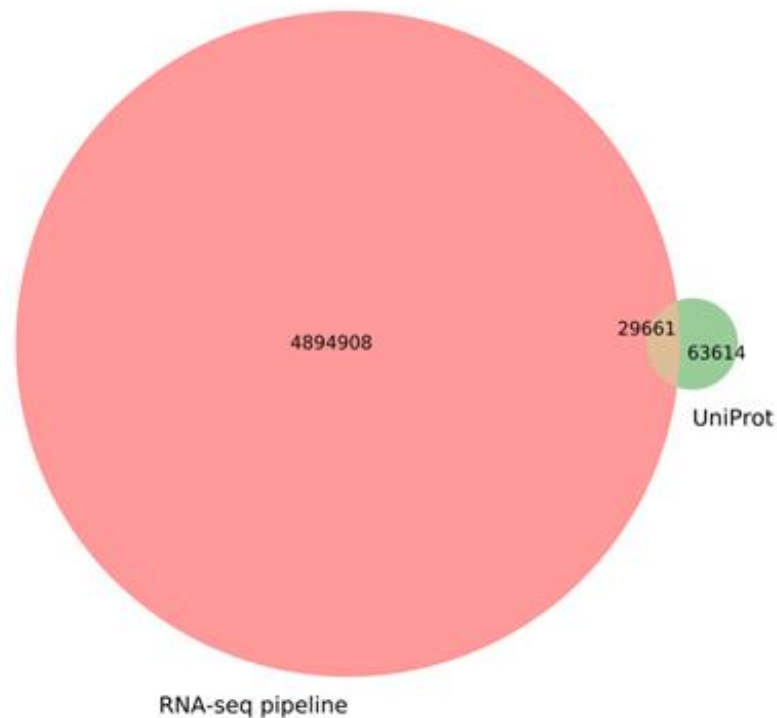
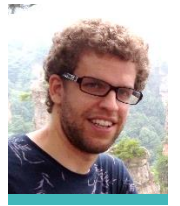


Nanopore RNA-seq:

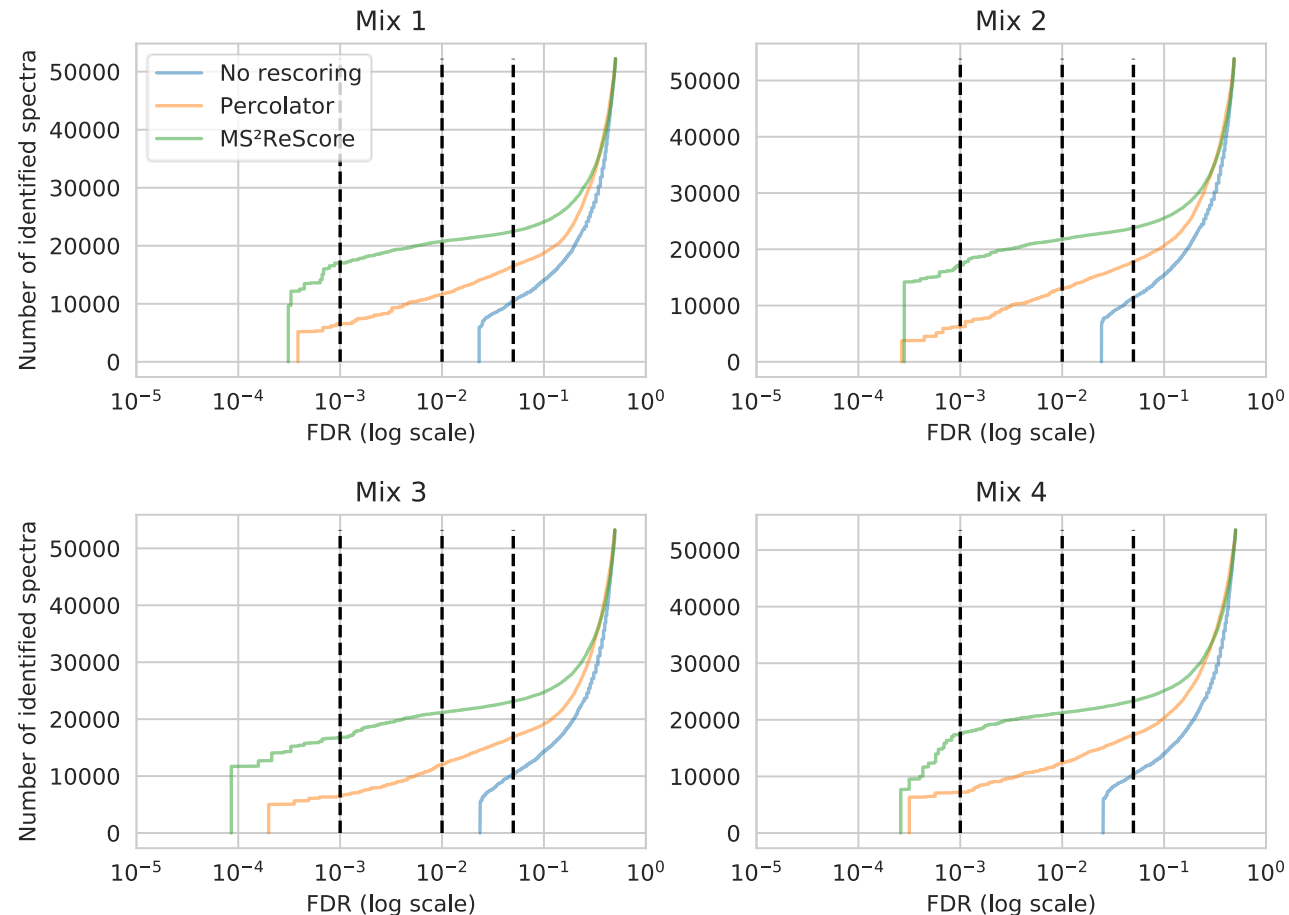
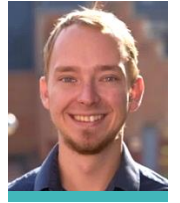
- Extended search space
- Human HCT116 cells
- HCD orbitrap acquisition



MS²Rescore in proteogenomics: Higher ID rate, at a 10-fold lower FDR threshold



MS²Rescore in metaproteomics: From 0 to 20 000 identified spectra

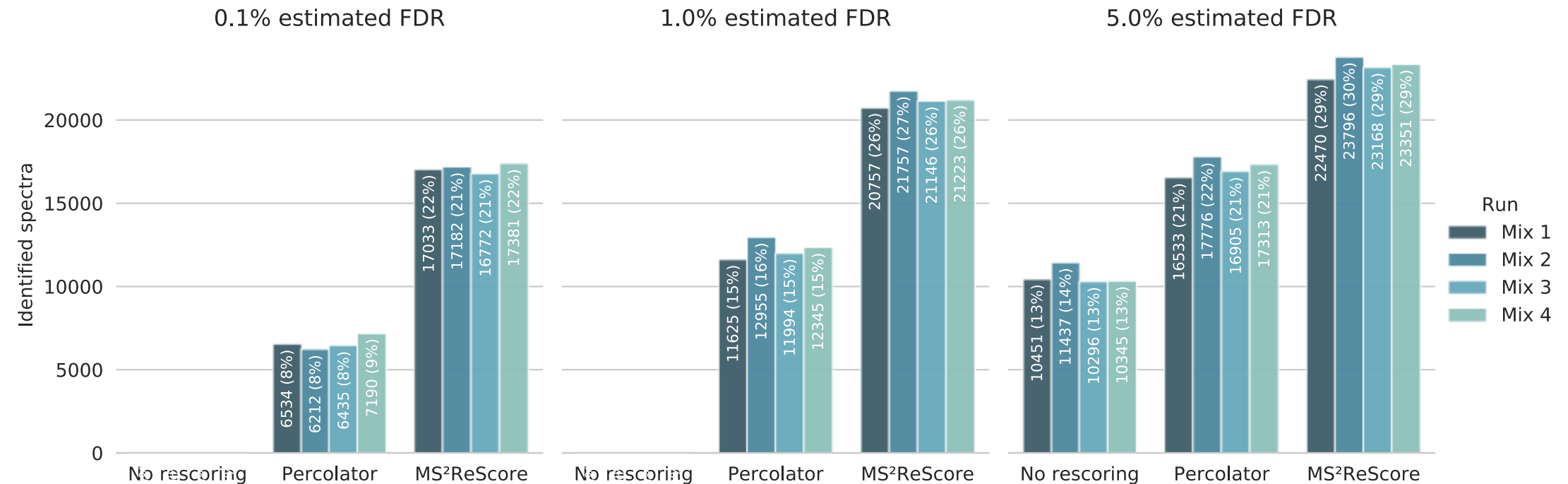
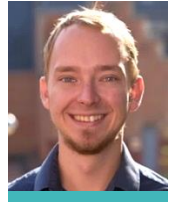


iPRG 2020 study:

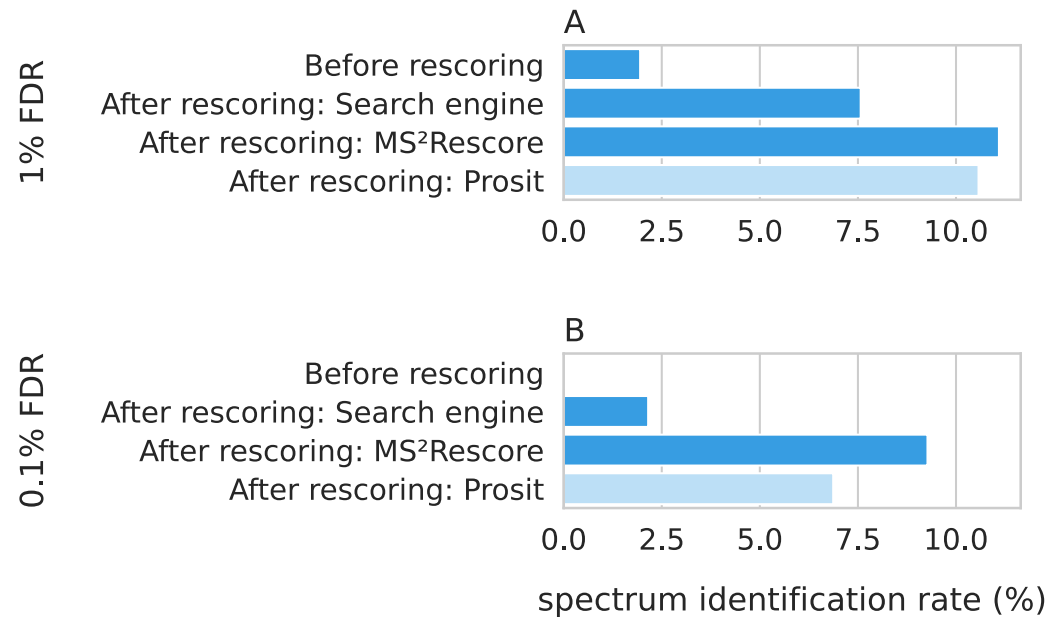
- 4 microbial mixes, with unknown sequences
- Extremely large search space
- CID ion trap acquisition



MS²Rescore in metaproteomics: From 0 to 20 000 identified spectra



MS²Rescore in immunopeptidomics: +46% spectrum identification rate

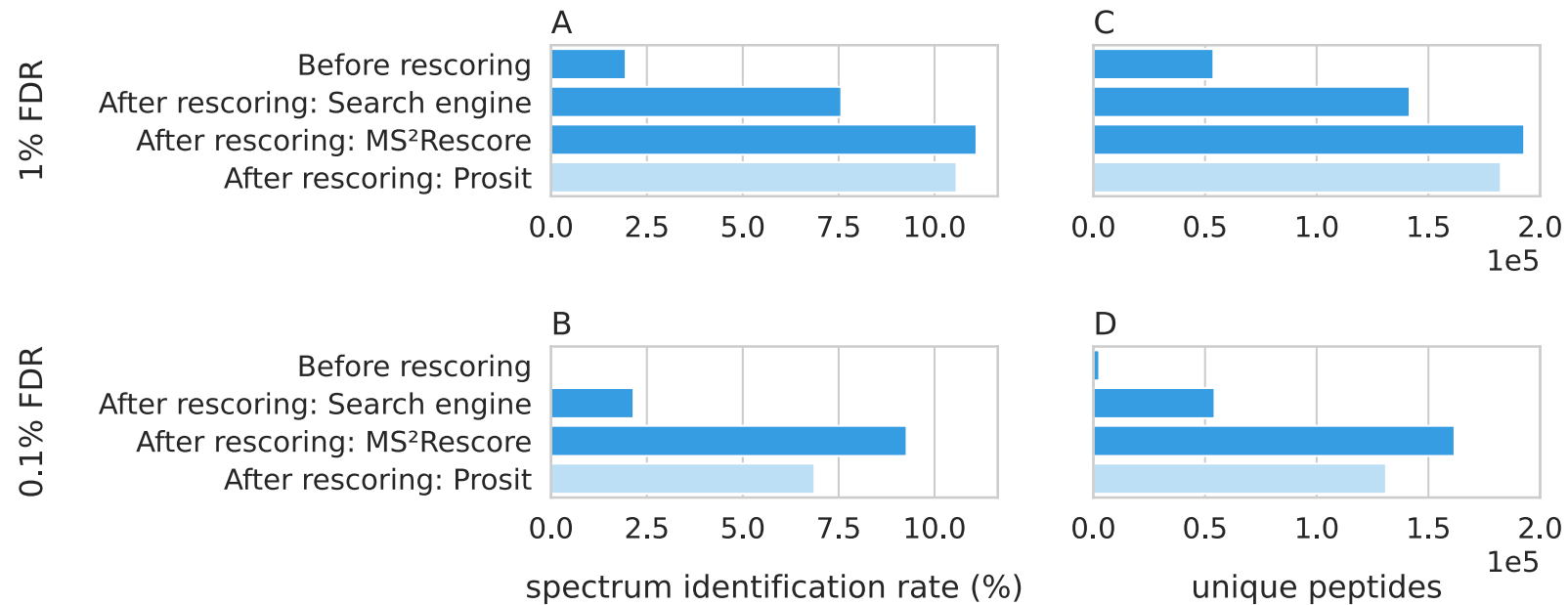


Immunopeptidomics:

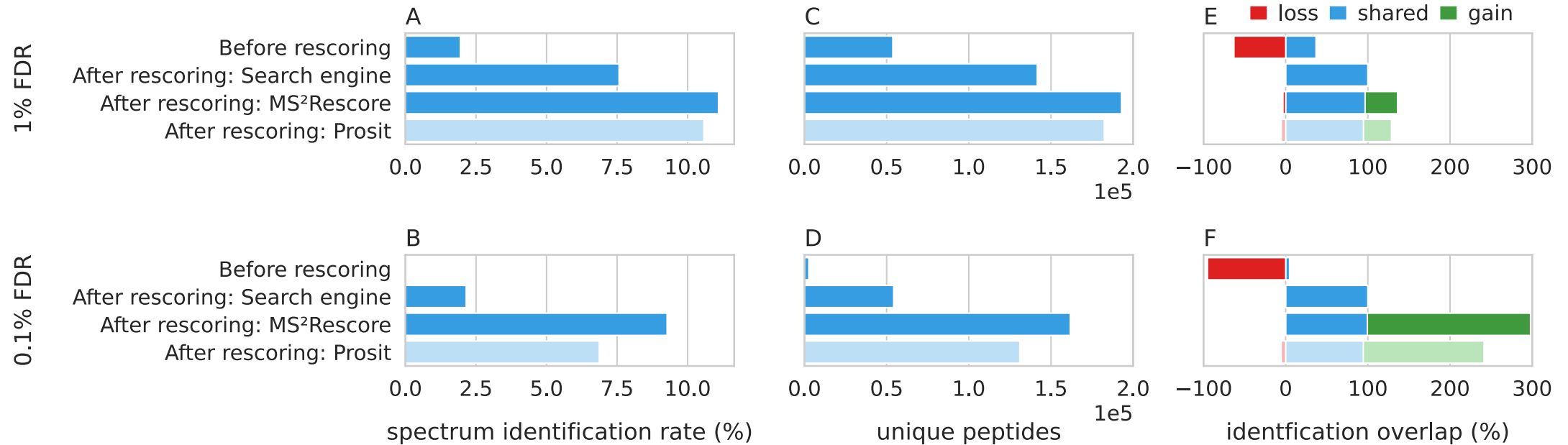
- non-specific cleavage
- peptides with low ionization and fragmentation efficiency



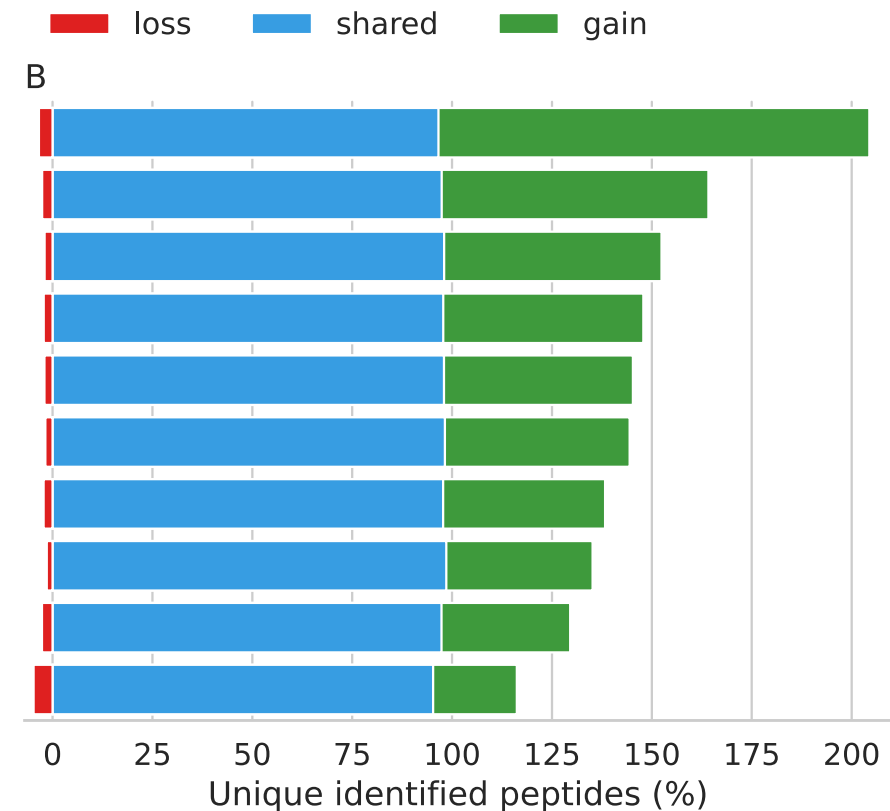
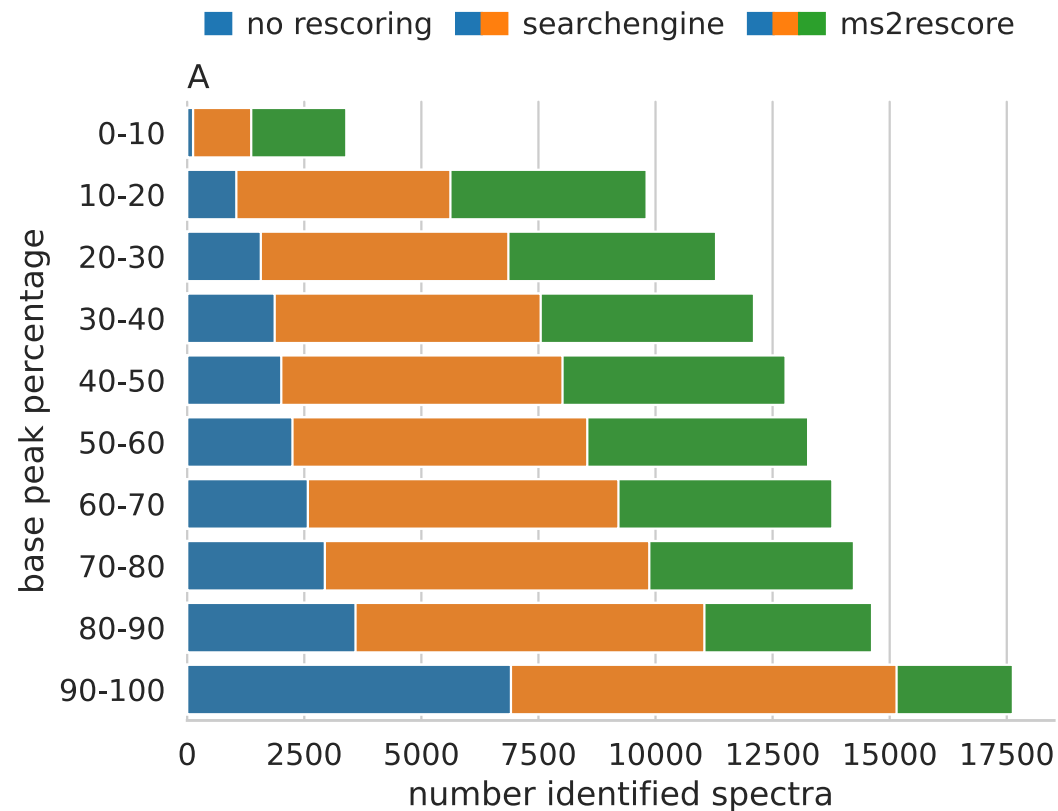
MS²Rescore in immunopeptidomics: +36% unique identified peptides



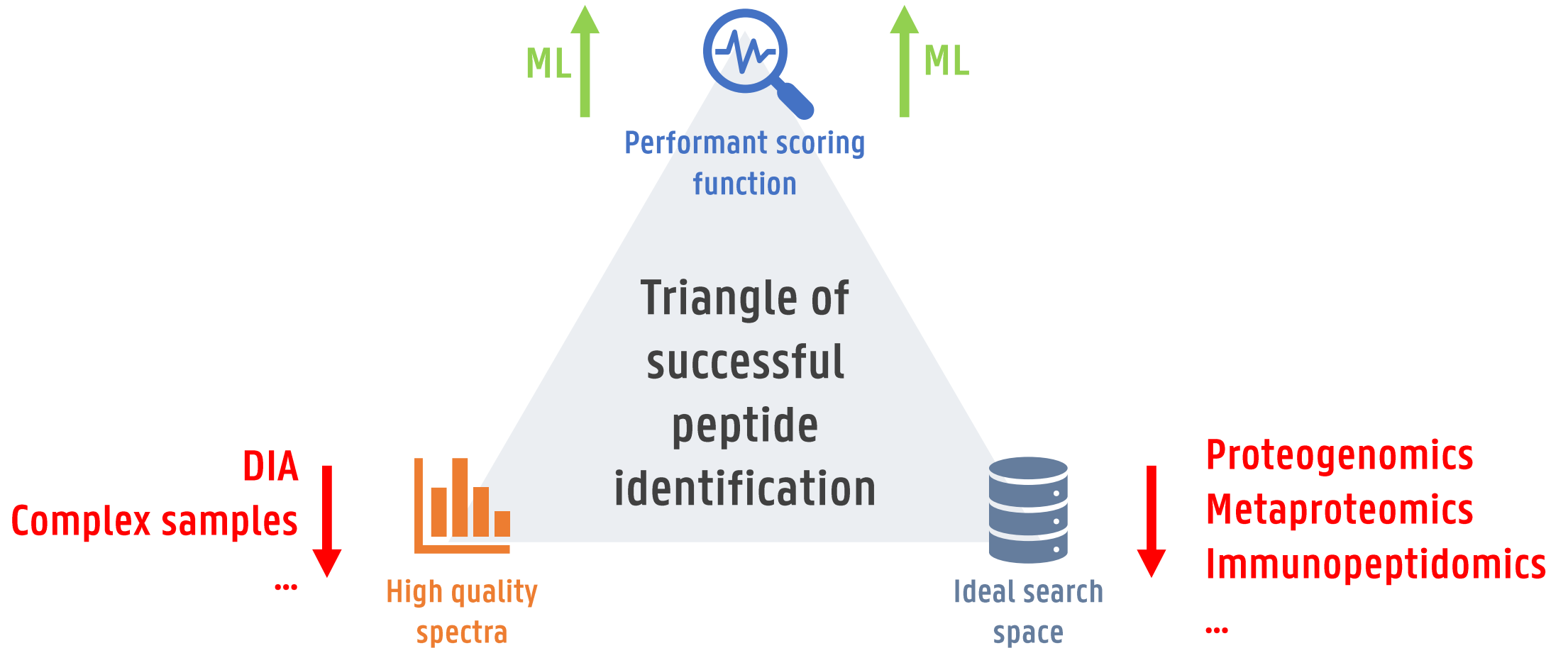
MS²Rescore in immunopeptidomics: +36% unique identified peptides



MS²Rescore in immunopeptidomics: Recovery of low-abundance peptides



In challenging experimental workflows, machine learning can rescue peptide identification rates



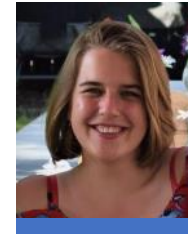
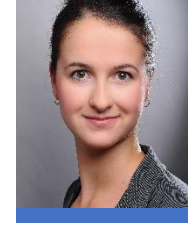
MS²Rescore is freely available on GitHub as a Python package, or as an easy-to-install GUI

The screenshot shows the GitHub repository for **compomics / ms2rescore**. The repository is public and has 8 branches and 9 tags. The main branch is **master**. The repository description is: "Sensitive PSM rescoring with predicted MS² peak intensities using MS²PIP, DeepLC, and Percolator." The repository has 6 stars and 5 forks. The repository is licensed under Apache-2.0. The repository has 5 contributors. The repository has 8 releases, with the latest release being **v2.1.2** on Apr 19. The repository has a README.md file. The repository has a setup.py file. The repository has a configuration.md file. The repository has a LICENSE file. The repository has a MANIFEST.in file. The repository has a .gitignore file. The repository has a tests directory. The repository has a ms2rescore directory. The repository has a gui-windows directory. The repository has an examples directory. The repository has a .github/workflows directory. The repository has a Merge pull request #70 from compomics/Fix/Comet-pin. The repository has a commit d37944f 5 days ago with 429 commits. The repository has a commit 7 months ago with the message "Add token to upload action". The repository has a commit 3 years ago with the message "Add files via upload". The repository has a commit 7 months ago with the message "Zip Windows installer scripts upon release". The repository has a commit 7 months ago with the message "Update README". The repository has a commit 5 days ago with the message "Set extract spectrum index default back to true". The repository has a commit 4 months ago with the message "Add more tests". The repository has a commit 8 months ago with the message "Add initial gui and pyinstaller support". The repository has a commit 4 years ago with the message "Create LICENSE". The repository has a commit 2 years ago with the message "Bugfixes:". The repository has a commit 2 months ago with the message "Update README.md". The repository has a commit 7 months ago with the message "Remove absolute maxquant mgf dir option". The repository has a commit 3 months ago with the message "Minor gui fixes".

The screenshot shows the MS²Rescore GUI configuration window. The window has a title bar with the MS²Rescore logo and a close button. The window has a header with the MS²Rescore logo and the text "MS²Rescore Sensitive PSM rescoring with MS²PIP, DeepLC, and Percolator." The window has a tabbed interface with three tabs: "General configuration", "MaxQuant settings", and "MS²PIP settings". The "General configuration" tab is selected. The "General configuration" section has the following fields:

- Identification file (required)**: Path to identification file (pin, mzid, msms.txt, tandem xml...).
- Spectrum file directory**: Path to MGF file or directory with MGF files (default: derived from identification file).
- Configuration file**: Path to MS²Rescore configuration file (see online documentation).
- Temporary file directory**: Path to directory to place temporary files.
- Output filename prefix**: Name for output files (default: derive from identification file).
- Select pipeline / search engine**: Identification file pipeline to use, depends on the search engine used. By default, this is inferred from the input file extension.
- Logging level**: Controls the amount information that is logged.

The window has a footer with two buttons: "Cancel" and "Start".



@compomics
www.compomics.com

@RalfGabriels



ENABLING NOVEL AND CHALLENGING PROTEOMICS WORKFLOWS WITH MS²RESCORE

Ralf Gabriels

EuPA ECR Day 2022

28/06/2022