

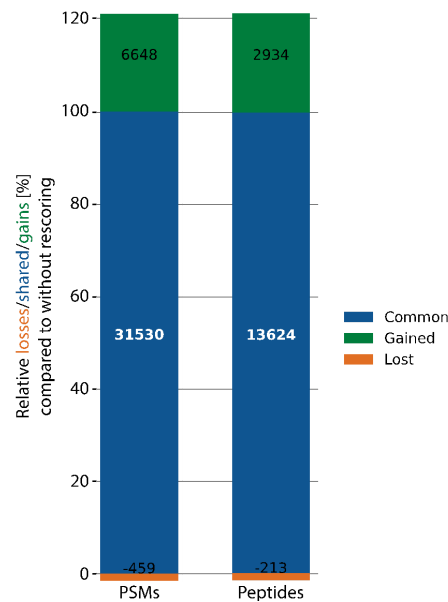


Figure 1: Stacked bar chart showing the number of confident PSMs (Peptide-Spectrum-Matches) (left) and peptides (right) below 1% FDR (False Discovery Rate) lost (orange), shared (blue) and gained (green) when rescoring using (Prosit) compared to the MSFragger [1] results.

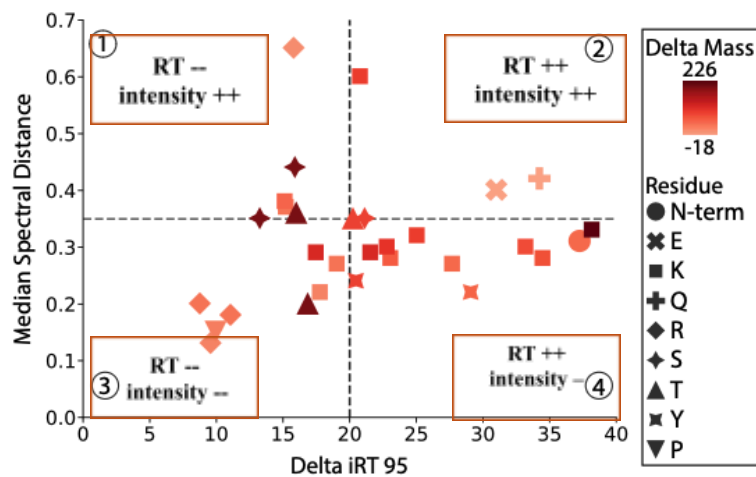


Figure 2: Impact of different PTMs on retention time and the fragment intensities when compared to its unmodified peptide counterparts. A spectral distance of 0.35 and delta iRT of 20 are commonly observed cutoffs in rescoring when differentiating correct from incorrect matches.

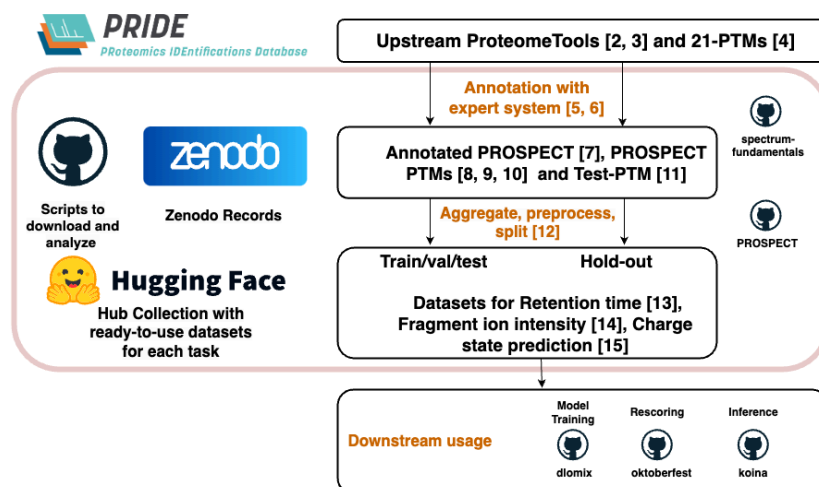


Figure 3: Lineage of the data, highlighting the contribution and resources provided in the scope of this work to facilitate accessibility of the datasets.

References

- [1] Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, 14(5), 513-520.
- [2] Daniel P Zolg and Kuster Bernhard. ProteomeTools. <https://www.ebi.ac.uk/pride/archive/projects/PXD009449>, 2018.
- [3] Daniel P Zolg and Kuster Bernhard. ProteomeTools. <https://www.ebi.ac.uk/pride/archive/projects/PXD023119>, 2021.
- [4] Daniel P Zolg and Kuster Bernhard. ProteomeTools. <https://www.ebi.ac.uk/pride/archive/projects/PXD023120>, 2021.
- [5] Nadin Neuhauser, Annette Michalski, Jürgen Cox, and Matthias Mann. Expert system for computer-assisted annotation of ms/ms spectra. *Molecular & Cellular Proteomics*, 11(11):1500–1509, 2012.
- [6] Wilhelm Lab. Spectrum fundamentals. https://github.com/wilhelm-lab/spectrum_fundamentals, 2023.
- [7] Omar Shouman, Wassim Gabriel, Victor-George Giurcoiu, Vitor Sternlicht, and Mathias Wilhelm. PROSPECT: Labeled tandem mass spectrometry dataset for machine learning in proteomics. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 32882–32896. Curran Associates, Inc., 2022.
- [8] Wassim Gabriel, Omar Shouman, and Mathias Wilhelm. PROSPECT PTMs Dataset - Multi-PTM. DOI:<https://doi.org/10.5281/zenodo.11472525>, 2024.
- [9] Wassim Gabriel, Omar Shouman, and Mathias Wilhelm. PROSPECT PTMs Dataset - TMT. DOI:<https://doi.org/10.5281/zenodo.8221499>, 2023.
- [10] Wassim Gabriel, Omar Shouman, and Mathias Wilhelm. PROSPECT PTMs Dataset - TMT-PTM. DOI:<https://doi.org/10.5281/zenodo.11474099>, 2024.
- [11] Wassim Gabriel, Omar Shouman, and Mathias Wilhelm. PROSPECT PTMs Dataset - Test-PTM. DOI:<https://doi.org/10.5281/zenodo.11477731>, 2024.
- [12] Omar Shouman, Wassim Gabriel, and Mathias Wilhelm. PROSPECT on GitHub. <https://github.com/wilhelm-lab/PROSPECT>, 2023.
- [13] Wilhelmlab - Computational Mass Spectrometry. prospect-ptms-irt (revision 82b3196), 2024.
- [14] Wilhelmlab - Computational Mass Spectrometry. prospect-ptms-ms2 (revision e8702d2), 2024.
- [15] Wilhelmlab - Computational Mass Spectrometry. prospect-ptms-charge (revision 48ab32c), 2024.
- [16] Omar Shouman, Wassim Gabriel, and Mathias Wilhelm. DLOmix on GitHub. <https://github.com/wilhelm-lab/dlomix>, 2024.
- [17] Mario Picciani, Wassim Gabriel, Victor-George Giurcoiu, Omar Shouman, Firas Hamood, Ludwig Lautenbacher, Cecilia Bang Jensen, Julian Müller, Mostafa Kalhor, Armin Soleymaniniya, et al. Oktoberfest: Open-source spectral library generation and rescoring pipeline based on prosit. *Proteomics*, 24(8):2300112, 2024.
- [18] Lautenbacher, L., Yang, K., Kockmann, T., Panse, C., Chambers, M., Kahl, E., ... & Wilhelm, M. (2024). Koina: Democratizing machine learning for proteomics research. *bioRxiv*, 2024-06.