

# quantms and ms2rescore: deep analysis of large-scale public proteomics datasets

Dai Chengxin<sup>1,2</sup>, Ralf Gabriels<sup>3</sup>, Robbin Bouwmeester<sup>3</sup>, Timo Sachsenberg<sup>4,5</sup>, and Yasset Perez-Riverol<sup>6</sup>

<sup>1</sup>State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, 102206, Beijing, China

<sup>2</sup>International Academy of Phronesis Medicine (Guangdong), 510320, Guangdong, China

<sup>3</sup>VIB-UGent Center for Medical Biotechnology, VIB, 9052 Ghent, Belgium

<sup>4</sup>Department of Computer Science, Applied Bioinformatics, University of Tübingen, Tübingen, Germany

<sup>5</sup>Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany

<sup>6</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, United Kingdom

## Abstract

The exponential growth of public proteomics datasets has outpaced the capacity of traditional desktop tools for large-scale automated analysis. This study presents an integrated workflow combining quantms, a cloud-native pipeline, with MS2Rescore, a machine learning-driven rescore tool, to enable deep reanalysis of massive proteomic datasets. Leveraging the Nextflow workflow engine for parallel computing, the pipeline integrates fragment ion intensity and retention time predictions from MS2PIP and DeepLC to optimize peptide-spectrum match reliability via Percolator. Across four benchmark datasets spanning label-free quantification, TMT labeling, immunopeptidomics, and phosphoproteomics, the workflow demonstrates a 16–22.8% increase in

identified spectra compared to traditional approaches, with hundreds of newly quantified proteins and phosphosites. The results highlight that multi-search-engine collaboration with machine learning-derived features not only enhances identification sensitivity but also deepens quantitative insights for downstream biological discoveries. This approach provides a reproducible solution for large-scale reanalysis of public proteomics data, aligning with FAIR principles for scientific accessibility and reuse.

**Keywords:** Proteomics, Reanalysis, Workflow, Machine learning.

## 1 Introduction

In recent years, the field of proteomics has experienced rapid growth in the availability of publicly accessible datasets, accompanied by a shift toward studies analyzing larger sample cohorts. As of December 2023, the PRIDE database contained over 25,000 public datasets, including a substantial increase in large-scale submissions comprising more than 100 instrument files—from 100 datasets in 2014 to 4,435 in 2024 [8]. Concurrently, transformative advances in proteomic data processing software have enabled deeper and more precise exploration of the proteome. However, conventional desktop tools such as MaxQuant [2], pFind [9], MSFragger [7], and Proteome Discoverer are limited in their capacity to perform automated, large-scale quantitative analyses in cloud or distributed environments, hindering the reanalysis of extensive experiments on standard workstations.

To address this limitation, we recently developed quantms, an open-source, cloud-based pipeline designed for massively parallel reanalysis of quantitative proteomic datasets [3]. The pipeline is highly modular and flexible, accommodating a wide range of quantitative proteomics approaches. quantms automatically distributes computations using the Nextflow workflow engine across one or more computing nodes, depending on the number of instrument files and samples [5]. To ensure traceability and reproducibility, the pipeline is built entirely on standardized open file formats and reproducible execution environments, adhering strictly

to the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles [10].

With the integration of machine learning (ML) into proteomics, various models have been developed to accurately predict peptide properties, such as MS2PIP [4] for fragment ion intensities and DeepLC [1] for retention time prediction. Early approaches employed decision trees and single-layer neural networks, while more recent deep learning models such as Prosit [6] achieve significantly improved accuracy for predicting fragment ion intensities and retention times. These highly accurate predictions enable superior matching of experimental data to theoretical expectations and have reinvigorated rescored strategies in proteomics.

Previously, quantms did not leverage measurable peptide properties such as fragment ion intensities and retention times. To overcome this limitation, we integrated MS2Rescore into quantms and incorporated customized features following Nextflow and nf-core best practices. We demonstrate that the enhanced pipeline supports in-depth analysis of large-scale public proteomics datasets across diverse experimental designs, including LFQ, TMT, immunopeptidomics, and phosphoproteomics studies.

## 2 Methods

### 2.1 MS/MS Data and quantms Search Results

To develop and evaluate the performance of the quantms-integrated MS2Rescore workflow, four publicly available benchmark datasets were downloaded from PRIDE Archive under the identifiers PXD001819, PXD019643, and PXD026824, and from CPTAC under the identifier PDC000125. The PXD001819 dataset contains 48 Sigma UPS1 proteins spiked into a background of yeast cell lysate at nine different concentrations: 0.05, 0.125, 0.25, 0.5, 2.5, 5, 12.5, 25, and 50 fmol/L to evaluate quantification performance. We evaluated five different quantms workflow settings: (1) Comet, (2) Comet and MSGF+, (3) Comet with MS2Rescore, (4) Comet and MSGF+ with MS2Rescore, and (5) Comet, MSGF+, and SAGE with MS2Rescore to explore multiple search engines and their integration with MS2Rescore

for improved identification and quantification results. All search parameters were the same as described in the previous publication. The quantms and MaxQuant search results at PSM and protein group levels are also available in parquet format in supplemental File S1.

## 2.2 Rescoring Features and Postprocessing

Search engine features are generated using the OpenMS toolkit. The quantms-rescoring Python package integrates MS2Rescore and calculates various meaningful features based on the DeepLC-predicted and observed retention times, and the MS2PIP-predicted and observed MS2 peak intensities. Signal-to-noise ratio (SNR) features are also calculated for specific cases in the quantms-rescoring Python package. The Percolator module was subsequently used to rescore the common pool of PSMs for different feature combinations using different strategies: (1) a baseline setting with only the search engine features, (2) the baseline setting combined with features from MS2Rescore, and (3) the baseline setting combined with features from MS2Rescore and the quantmsutils package.

# 3 Results

## 3.1 quantms with MS2Rescore enhancing identification and quantification in label-free experiments

To systematically evaluate the performance of the quantms-integrated MS2Rescore workflow at both identification and quantification levels, we first analyzed the public benchmark dataset PXD001819. For identification benchmarking, five different identification workflows were designed and compared, including (1) Comet with Percolator, (2) two search engines with Percolator, (3) Comet with MS2Rescore features and Percolator, (4) two search engines with MS2Rescore features and Percolator, and (5) three search engines with MS2Rescore features

and Percolator to determine whether features from intensity-based and retention time-based predictors enhanced the identification and quantification process.

Filtering for a set of significant PSMs is generally performed using q-values and the resulting FDR estimation. The most valuable measure to compare the sets with and without additional features is therefore the FDR, as this shows the actual yield at PSM-level stringency. Figure 1 clearly shows that consensus scores from two search engines substantially improve identification rates at the same FDR level. The multiple search engines consensus score improved the number of identified spectra by 17% compared with Comet identification alone. The number of identified spectra increased by a further 16% after adding MS2PIP and DeepLC features. Feature calculation was performed using the standalone MS2Rescore tool. We further explored the effect of different settings on protein quantification results, especially for adding extra features. More low-abundance UPS1 proteins were quantified after adding MS2Rescore features (Figure 2B).

Moreover, we extracted the top 20 weights from Percolator to explore the impact of different features on rescore for different search engines, shown in Supplemental Figure 1. We can see that more than half of the features are from MS2Rescore, and the distributions of features are different for the three search engines. The weight of SpecPearsonNorm is positive, indicating that a high SpecPearsonNorm value corresponds to a better hit. The weight of RtDiffBest is negative, indicating that large differences between observed and predicted retention time correspond to a worse score. It is interesting that peptide length becomes a significant feature after adding MS2Rescore features. The positive value demonstrates higher scores for longer peptides. One potential explanation for this phenomenon is that short peptides are less conducive to search engine identification due to fewer fragment ions.

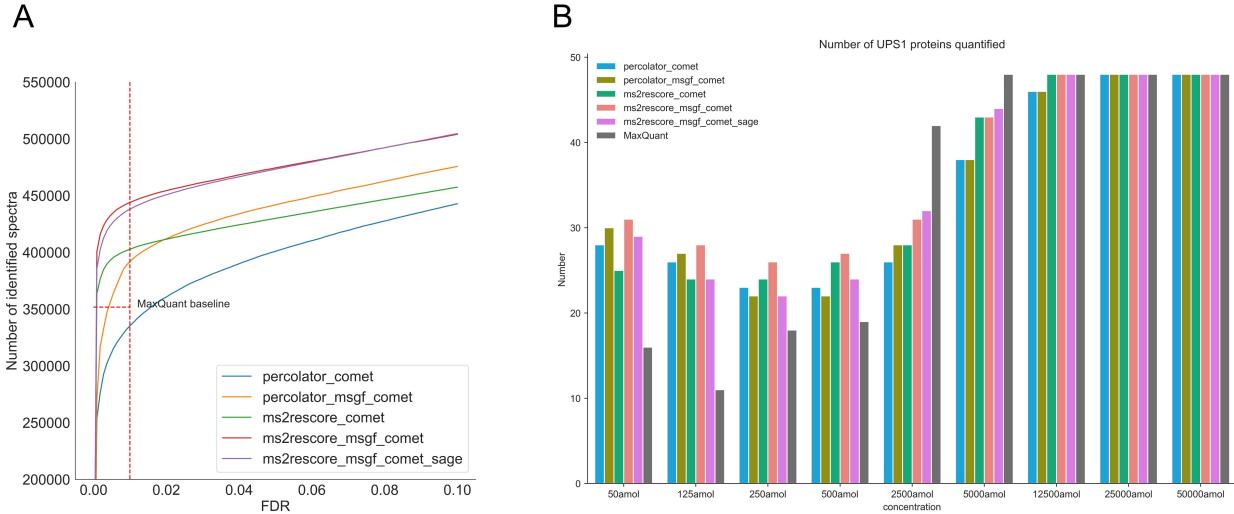


Figure 1: The number of identified spectra as a function of differing FDR levels for different workflow settings. (A) Results for the PXD001819 dataset. (B) Results for the PXD001819 dataset showing protein quantification. Five different workflow settings are shown: the combination of Comet search engine with Percolator rescore, the combination of Comet with MSGF+, the combination of two search engines with MS2Rescore features, the combination of two search engines with MS2Rescore and SNR features, and the combination of three search engines with MS2Rescore and SNR features. Multiple search engines consensus identification, additional intensity-based and retention time-based features, and SNR features lead to both higher identification rates and high-quality identification.

### 3.2 quantms integrated with MS2Rescore drastically improves identification and quantification for large-scale TMT experiments

We then applied this workflow to large-scale TMT experiments from CPTAC. Multiple search engines and rescore significantly improved peptide identification and protein quantification, as shown in Figure 2. The PSM identification rate was improved by 3.6%, and 921 proteins were newly quantified when DeepLC and MS2PIP features were added, compared to the workflow without MS2Rescore enabled (Figure 2A,B). There are 59 newly quantified proteins whose abundance is located in the top 10% (Figure 2C). To further explore the impact of these newly quantified proteins from the two search engines with MS2Rescore features workflow on downstream analysis, protein differential expression results from this workflow are shown in Figure 2. There are 27 newly quantified proteins that are significantly

differentially expressed (Figure 2D). Among them, Yang et al. reported an association between significant upregulation of FOXG1 protein and prognosis in Clear Cell Renal Cell Carcinoma [10.18632/aging.204448]. This further illustrates that by incorporating rescoring features based on MS2PIP and DeepLC, the approach not only improves the sensitivity of spectral identification but also facilitates protein quantification, which ultimately contributes to biological knowledge discovery.

Moreover, we extracted the top 20 SVM weights from Percolator to explore the impact of different features on rescoring for different search engines, shown in Supplemental Figure S2. For Comet identification results, SpecPearsonNorm, DotProdIonYNorm, and RtDiffBest features have significant weights for final rescoring. The weight of SpecPearsonNorm is positive, indicating that a high SpecPearsonNorm value between MS2PIP prediction and experimental intensities corresponds to a better hit. The weight of RtDiffBest is negative, indicating that large differences between observed and calculated retention time correspond to a worse score. It is interesting that peptide length becomes a significant feature after adding MS2Rescore features. The positive value demonstrates higher scores for longer peptides. One potential explanation for this phenomenon is that short peptides are less conducive to search engine identification due to fewer fragment ions. For the other two search engines, similar distributions of weights can be observed. SpecPearsonNorm, DotProdIonYNorm, and RtDiffBest features still have significant weights.

### 3.3 Evaluation of quantms integrated with MS2Rescore on HLA

#### Class I Peptides

Furthermore, we evaluated the performance of the quantms identification workflow on the immunopeptides dataset PXD019643. Therefore, four different workflow settings were compared: the combination of Comet search engine with Percolator rescoring, the combination of Comet with MSGF+, the combination of two search engines with MS2Rescore features, and the combination of two search engines with MS2Rescore and SNR features. These

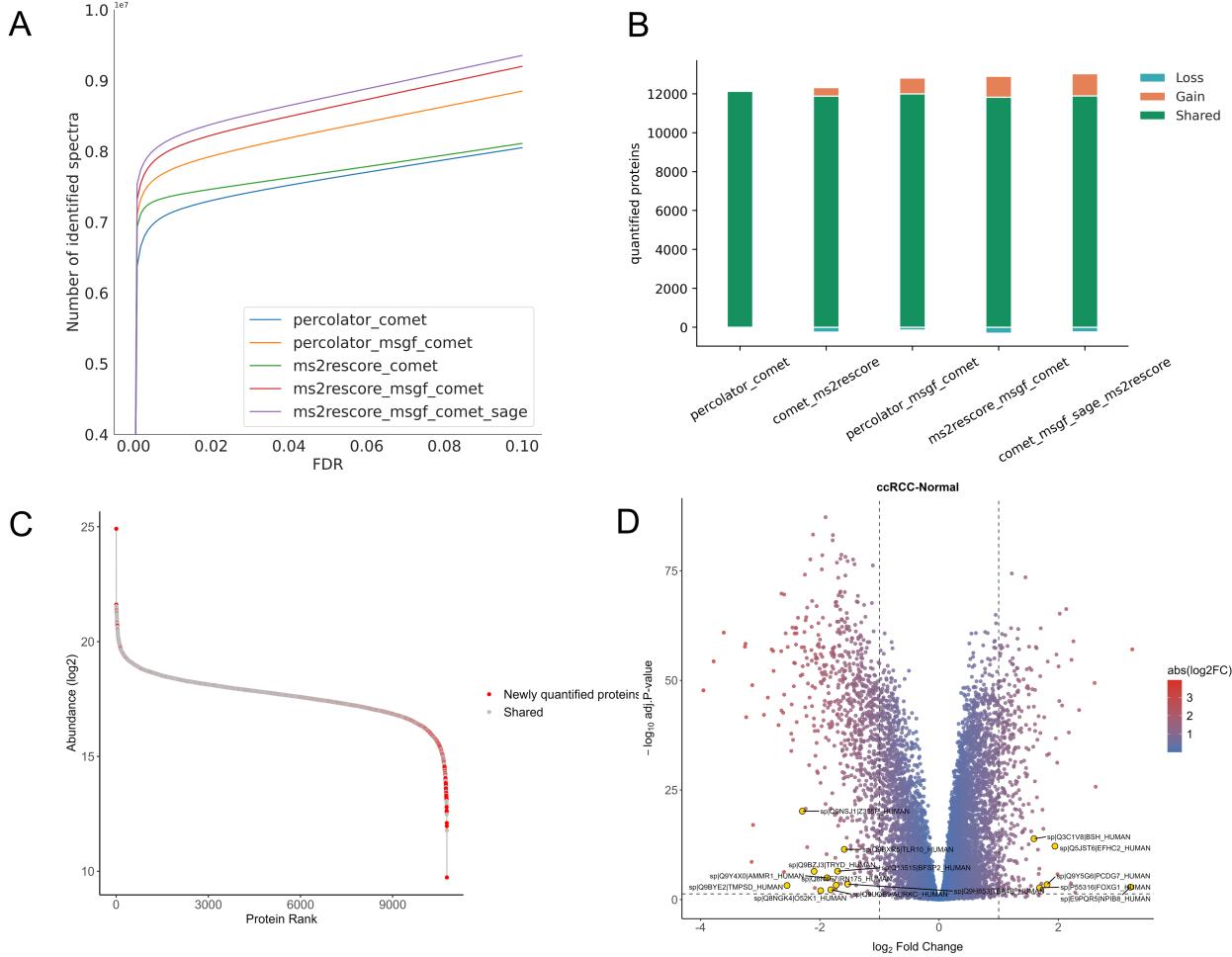


Figure 2: Comparison of identification and quantification results for different workflow settings. (A) The number of identified spectra as a function of differing FDR levels for different workflow settings. (B) The number of quantified proteins. The green part indicates the intersection between a workflow and Comet results. (C) The rank of protein abundance from Comet and MSGF+ with MS2Rescore. The red dots represent proteins quantified only in the Comet and MSGF+ with MS2Rescore workflow compared to the Comet and MSGF+ workflow without MS2Rescore. (D) The volcano plot for differential expression analysis from MSstatsTMT. The red dots represent proteins quantified only in the Comet and MSGF+ with MS2Rescore workflow compared to the Comet and MSGF+ workflow without MS2Rescore.

were compared in terms of the total amount of identifications as well as the number of unique identifications based on sequence. Overall, multiple search engines and rescoring with MS2Rescore substantially improved the spectrum identification rate compared to using only the Comet search engine or no rescore at both 1% and 0.1% FDR. Multiple search

engines achieved an 11.7% increase in the number of identified spectra compared to using only Comet search engines, and MS2Rescore features further increased the number of identified spectra by 22.8% compared to multiple search engines without MS2Rescore features, as shown in Figure 3.

The power of providing these predictions to Percolator is further illustrated when visualizing the distributions for decoy PSMs, rejected target PSMs, and accepted target PSMs from individual support and consensus supports (Figure 3C, D). The accepted target PSMs are clearly separable from the decoy and rejected target PSMs using only the Pearson correlation coefficient (PCC). The distributions for accepted targets from individual support and consensus supports are highly similar in that they both exhibit low retention time errors and high PCC values. This further demonstrates that quantms is effective for integrating the identification scores of multiple search engines.

### 3.4 quantms integrated with MS2Rescore facilitates the identification of phosphorylated peptides

In addition, we investigated the performance of our quantms workflow with MS2Rescore on post-translational modification experiments (PXD026824). For phosphoproteomics analyses, different workflow settings were evaluated: (1) Comet alone, (2) Comet combined with MSGF+, and (3) multiple search engines combined with MS2Rescore features. These were processed using Percolator to determine whether multiple search engines and MS2Rescore facilitated the identification and quantification process. In Figure 4, it is clearly visible that the introduction of multiple search engines consensus identification and MS2Rescore features extensively improves the stringency with which Percolator can classify PSMs. This is illustrated by the remarkable shift to the left compared with Comet alone. In addition, there is also a gain in identification rate, which is shown in the vertical direction of this figure. The two search engines consensus identification combined with MS2Rescore features enabled a 19% increase in the number of identified spectra compared to consensus identifi-

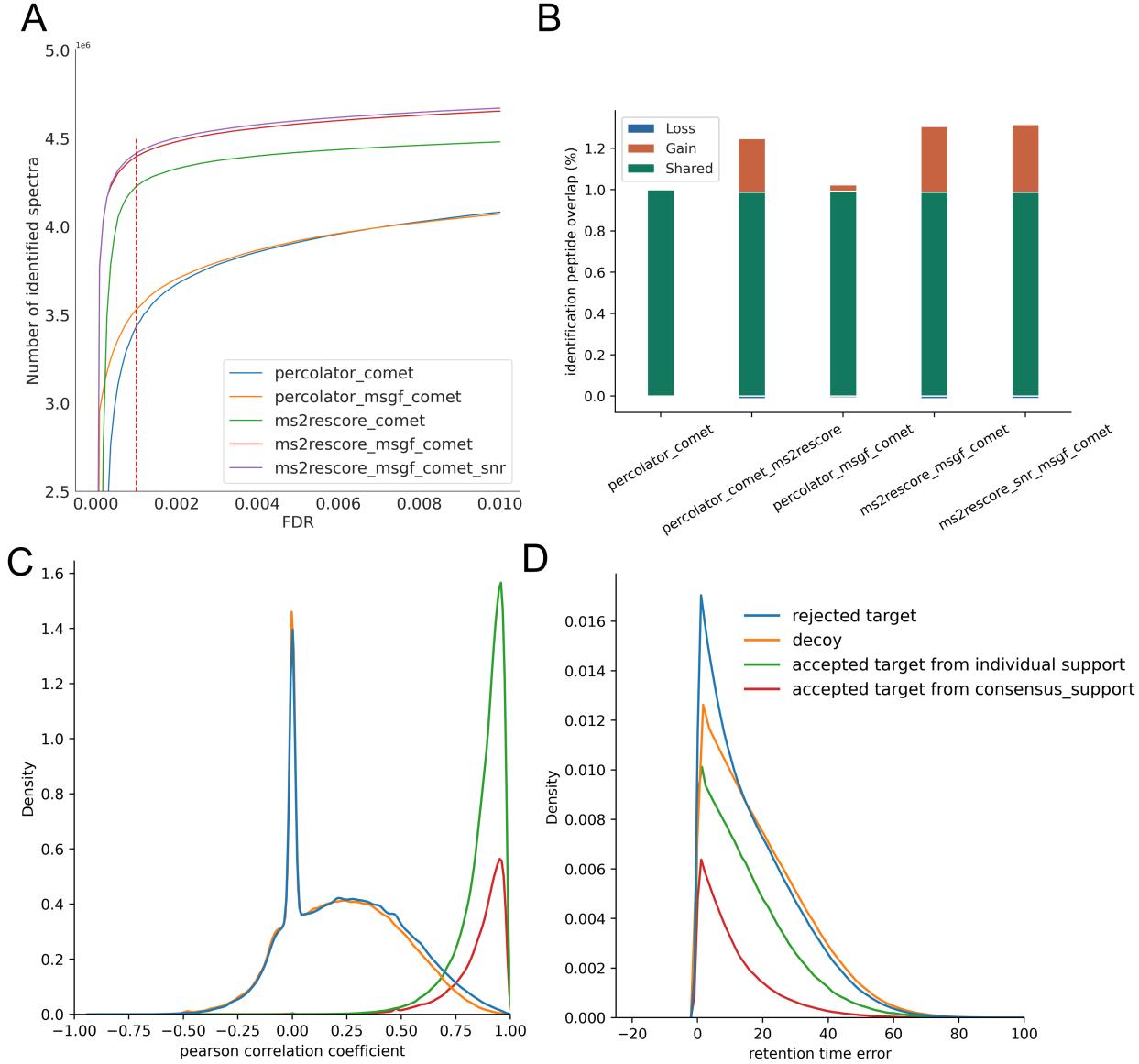


Figure 3: Comparison of immunopeptides identification results for different workflow settings. (A) Line plot showing the number of spectra identified for four workflow settings at different PSM FDR levels. (B) Percentage of unique identified peptides using different workflow settings. Density plots showing the distribution of the smallest retention time error between observed and predicted retention time (C) and the Pearson correlation between observed and predicted peak intensities (D) for each PSM split into decoys (red), rejected targets with q-value  $\geq 0.01$  (blue), and accepted targets from individual support and consensus support with q-value  $\geq 0.01$  (green). Note that the rejected targets distribution coincides with the decoy distribution.

cation without MS2Rescore (Figure 4A). The top 20 weights from Percolator are shown in Supplemental Figure S4. We can observe the same features as in the above results, such as

the DotProdIonYNorm feature.

Protein phosphosite false localization rate (FLR) control is necessary for phosphoproteomics analysis. Therefore, we also investigated the effect of different workflow settings on phosphorylated peptides at different FLR levels (Figure 4B, C). The search engines consensus identification combined with MS2Rescore features increased the number of phosphorylated peptide identifications by 17% at 0.01 local FLR compared to the approach without MS2Rescore. There are 1,345 phosphorylated peptides quantified only in the consensus identification combined with MS2Rescore process. Considering protein phosphosites, consensus identification combined with the MS2Rescore process newly reports 350 protein phosphorylation sites (Figure 4D). Altogether, these results show that quantms integrated with MS2Rescore can boost performance in phosphoproteomics.

## 4 Discussion

Advancements in deep learning-based tools have significantly improved the accuracy and sensitivity of peptide-spectrum match (PSM) rescoring, yet their integration into streamlined, reproducible, and quantitative proteomics workflows has remained limited, especially for large-scale public data analysis. In this study, we demonstrate the systematic incorporation of MS2Rescore into the open-source quantms pipeline and evaluate its impact across various experimental settings, including label-free quantification (LFQ), tandem mass tag (TMT)-based quantification, immunopeptidomics, and phosphoproteomics. Our results show that enhanced feature sets derived from MS2PIP and DeepLC not only improve identification rates but also translate into gains in quantification depth. Importantly, we provide evidence that the improved identifications via rescoring propagate downstream into quantification, leading to increased numbers of proteins with reliable abundance estimates, as well as enhanced detection of differentially expressed proteins.

To demonstrate the applicability and advantages of quantms integrated with MS2Rescore

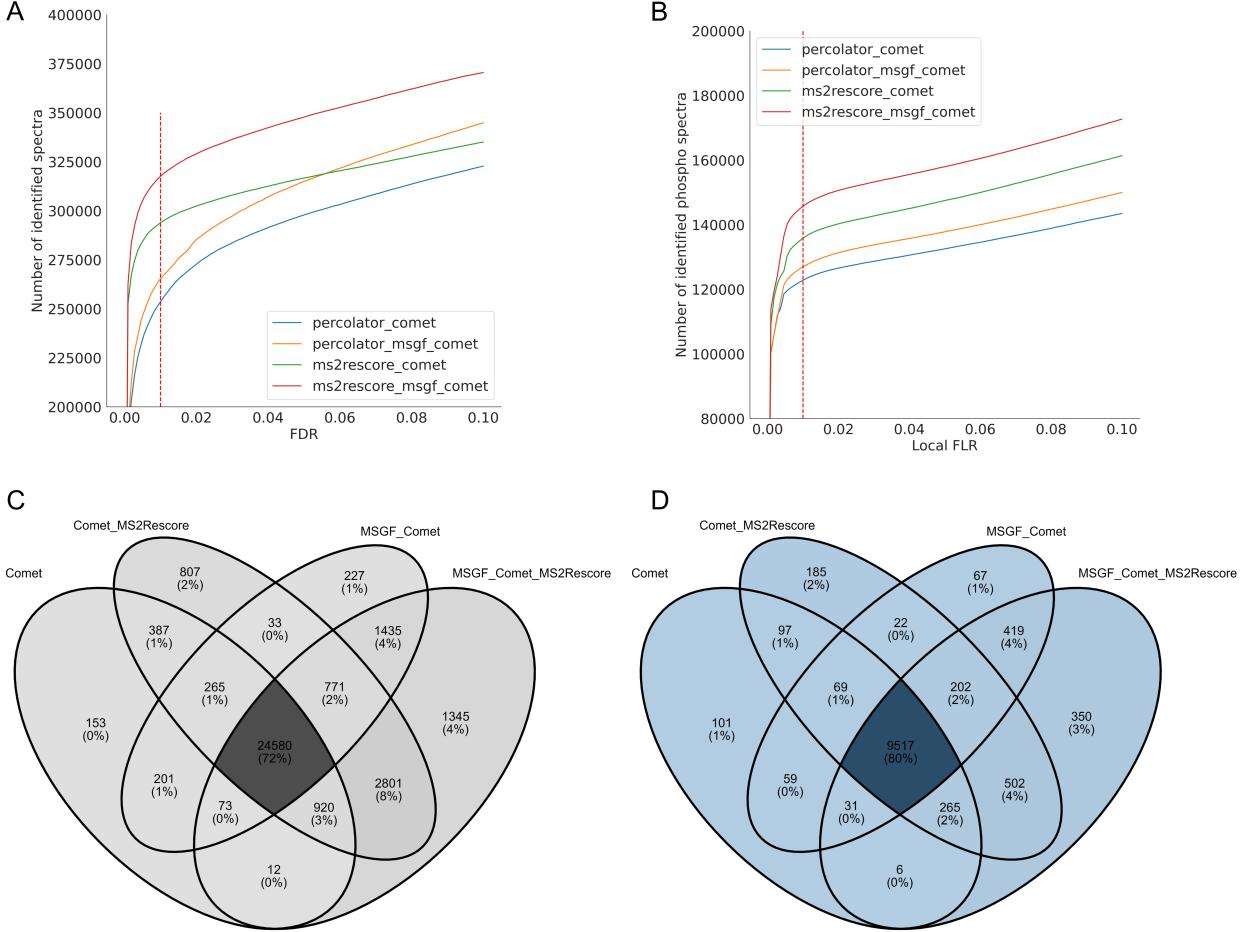


Figure 4: Comparison of phosphorylated peptide identification results for different workflow settings. (A) Line plot showing the number of spectra identified for four workflow settings at different PSM FDR levels. (B) Line plot showing the number of phosphorylated spectra identified for four workflow settings at different local false localization rate levels after an FDR of less than 0.01 at the PSM level. (C) Venn diagram of peptides quantified for four settings. (D) Venn diagram of protein phosphosites at protein FDR 0.01 and FLR 0.01 for four settings.

in quantitative proteomics, we performed a series of reanalyses using publicly available large-scale datasets with well-established benchmarks. In label-free experiments, MS2Rescore-enhanced workflows achieved significant increases in PSM and peptide identification while maintaining or improving quantification reproducibility across replicates. For TMT experiments, the integration of MS2Rescore into quantms increased the number of quantifiable proteins. In more challenging applications, such as immunopeptidomics and phosphoproteomics, the deep learning-based rescore pipeline consistently yielded higher identification

rates and improved sensitivity. These improvements contribute meaningful biological insights, such as phosphosite discovery and the identification of differentially expressed proteins with potential clinical relevance.

Overall, the results in this publication highlight the synergistic benefits of integrating machine learning-based features within end-to-end, cloud-based proteomics workflows. By embedding MS2Rescore into the quantms pipeline, we provide a practical and accessible solution for the community to leverage cutting-edge spectral prediction and retention time modeling without the need for complex manual configuration. This integrated approach enables reproducible reanalysis of large public datasets, increases identification confidence, and strengthens quantitative conclusions across diverse experimental designs. Future work will focus on expanding this approach to other types of post-translational modifications and exploring the integration of additional machine learning tools to further enhance proteomics data analysis capabilities.

## Acknowledgments

We thank all those who supported this research, including funding bodies and the proteomics community for making public datasets available. We also acknowledge the developers of the open-source tools used in this study, including MS2Rescore, MS2PIP, DeepLC, and Percolator, whose contributions made this work possible.

## References

- [1] R. Bouwmeester, R. Gabriels, N. Hulstaert, L. Martens, and S. Degroeve. DeepLC can predict retention times for peptides that carry as-yet unseen modifications. *Nature Methods*, 18(11):1363–1369, Nov. 2021.
- [2] J. Cox and M. Mann. MaxQuant enables high peptide identification rates, individu-

- alized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12):1367–1372, Dec. 2008.
- [3] C. Dai, J. Pfeuffer, H. Wang, P. Zheng, L. Käll, T. Sachsenberg, V. Demichev, M. Bai, O. Kohlbacher, and Y. Perez-Riverol. quantms: a cloud-based pipeline for quantitative proteomics enables the reanalysis of public proteomics data. *Nature Methods*, July 2024.
- [4] S. Degroeve and L. Martens. MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics (Oxford, England)*, 29(24):3199–3203, Dec. 2013.
- [5] P. Di Tommaso, M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C. Notredame. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319, Apr. 2017.
- [6] S. Gessulat, T. Schmidt, D. P. Zolg, P. Samaras, K. Schnatbaum, J. Zerweck, T. Knaute, J. Rechenberger, B. Delanghe, A. Huhmer, U. Reimer, H.-C. Ehrlich, S. Aiche, B. Kuster, and M. Wilhelm. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, 16(6):509–518, June 2019.
- [7] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, and A. I. Nesvizhskii. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods*, 14(5):513–520, May 2017.
- [8] Y. Perez-Riverol, C. Bandla, D. Kundu, S. Kamatchinathan, J. Bai, S. Hewapathirana, N. John, A. Prakash, M. Walzer, S. Wang, and J. Vizcaíno. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Research*, 53(D1):D543–D553, Jan. 2025.
- [9] L. Wang, D. Li, Y. Fu, H. Wang, J. Zhang, Z. Yuan, R. Sun, R. Zeng, S. He, and W. Gao. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21(18):2985–2991, Sept. 2007.

- [10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, Mar. 2016.

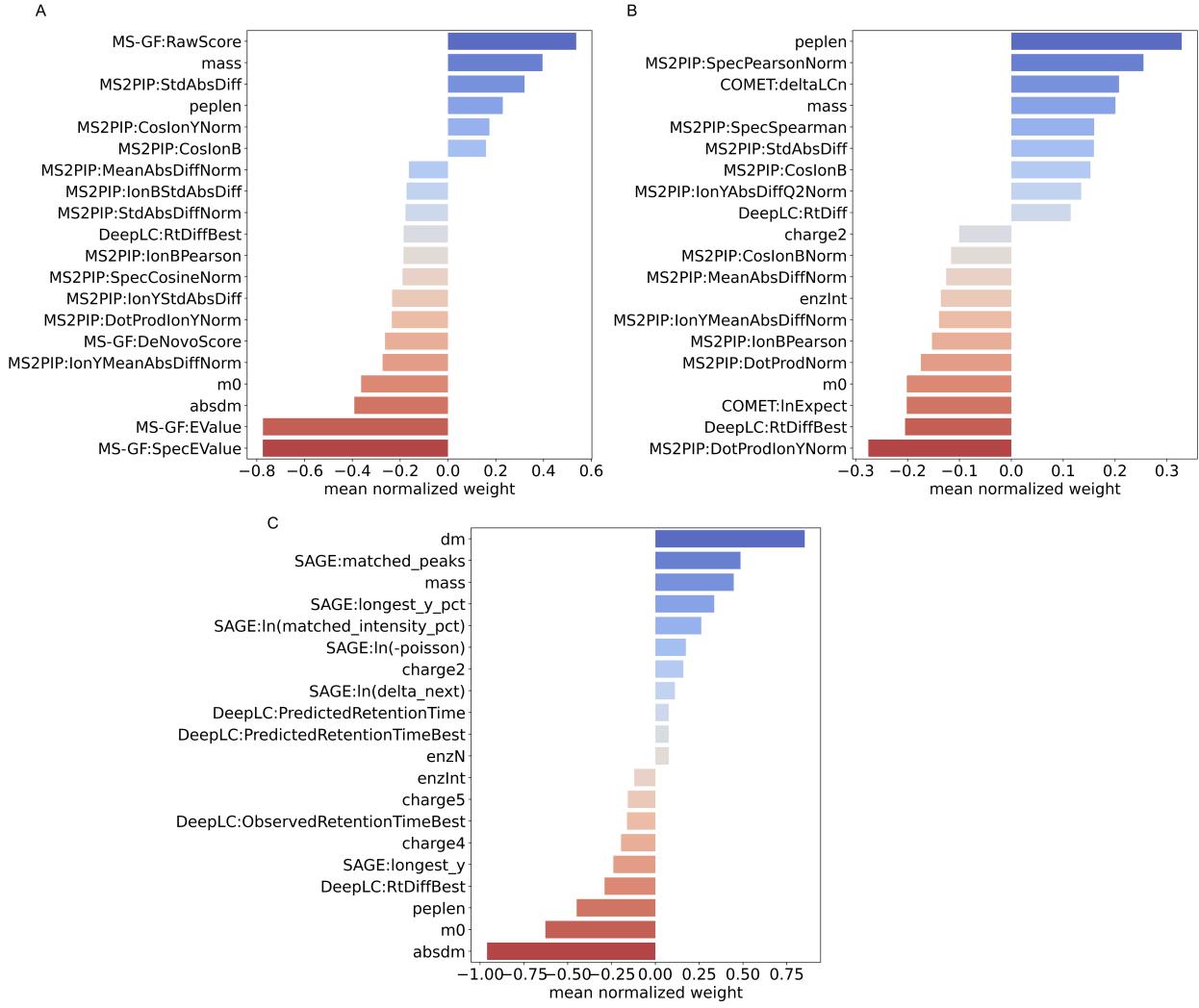


Figure S1: The weights that Percolator assigns to different features of the search engine feature vector in MSGF+ (A), Comet (B), and SAGE (C) combined with Percolator settings from PXD001819. The more different from zero a weight is (both positive and negative), the more important that feature is in the final Percolator classification.

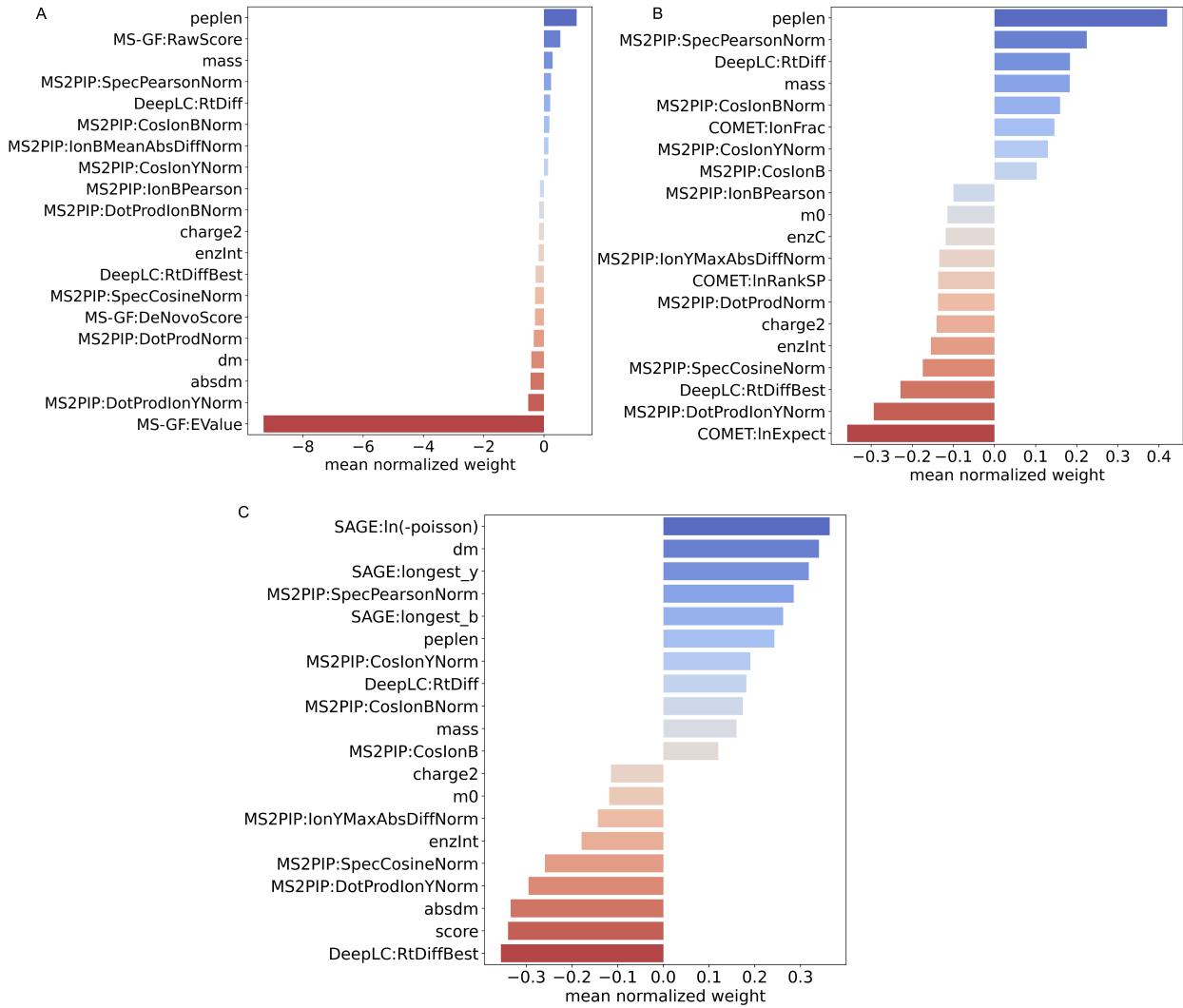


Figure S2: The top 20 normalized weights from Percolator for MSGF+ (A), Comet (B), and SAGE (C) identification results from PDC000125.

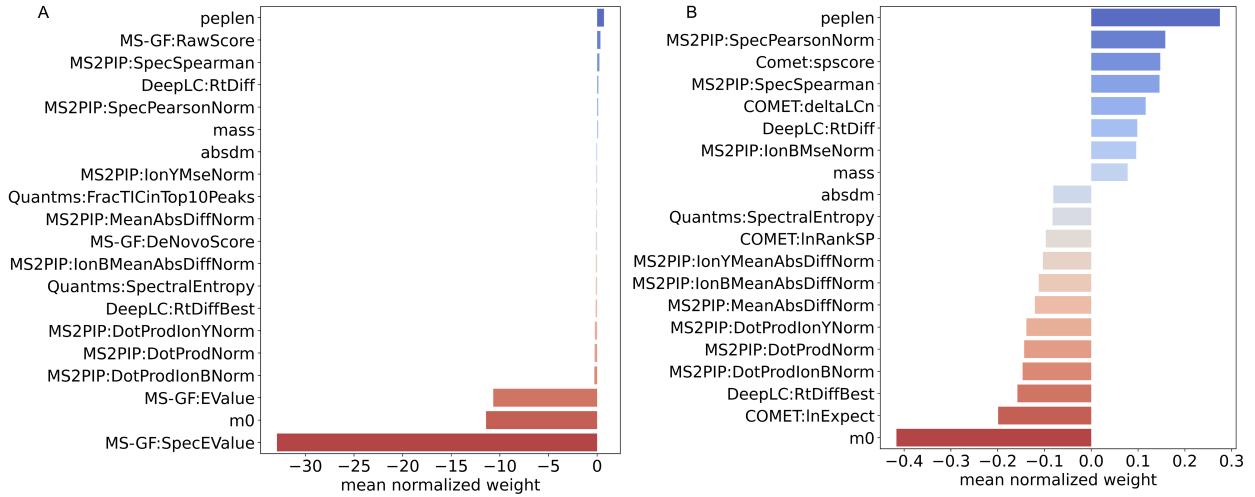


Figure S3: The top 20 normalized weights from Percolator for MSGF+ (A) and Comet (B) identification results in the PXD019643 dataset.

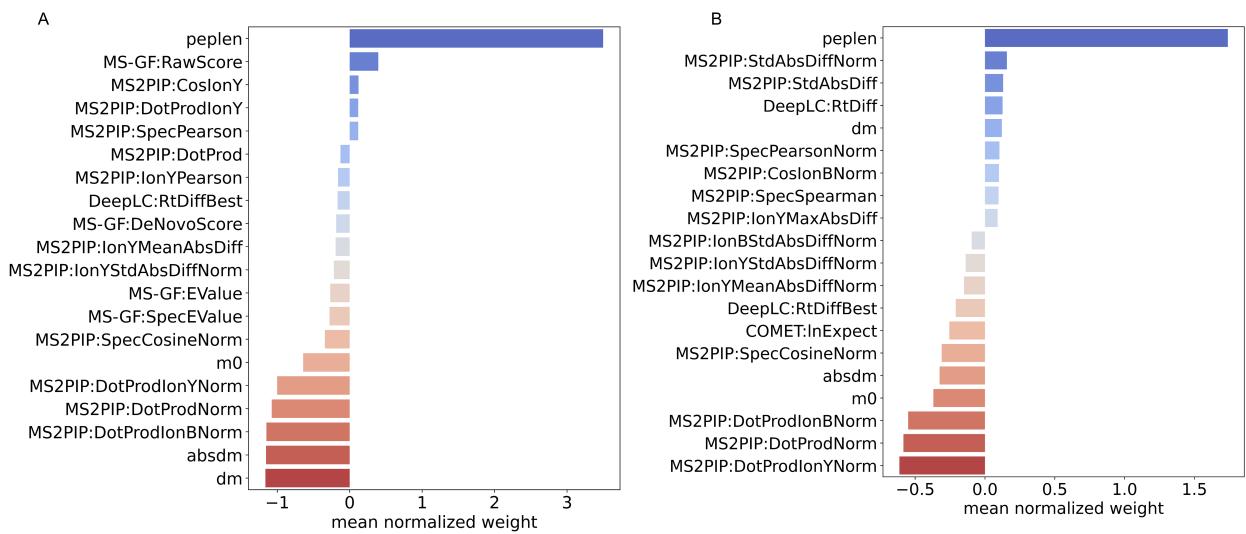


Figure S4: The top 20 normalized weights from Percolator for MSGF+ (A) and Comet (B) identification results in the PXD026824 dataset.