WOMBAT-P: benchmarking label-free proteomics data analysis workflows

Bouyssié, D.; Altiner, P.; Capella-Gutierrez, S.; Fernández, J.M.; Hagemeijer, Y.P.; Horvatovich, P.; ... ; Schwämmle, V.

Journal of proteome research

pubs.acs.org/jpr

Article

WOMBAT-P: Benchmarking Label-Free Proteomics Data Analysis Workflows

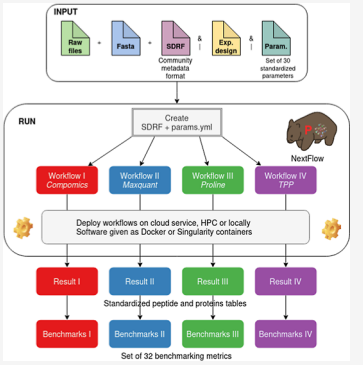
David Bouyssié, Pınar Altıner, Salvador Capella-Gutierrez, José M. Fernández, Yanick Paco Hagemeijer, Peter Horvatovich, Martin Hubálek, Fredrik Levander, Pierluigi Mauri, Magnus Palmblad, Wolfgang Raffelsberger, Laura Rodríguez-Navas, Dario Di Silvestre, Balázs Tibor Kunkli, Julian Uszkoreit, Yves Vandenbrouck, Juan Antonio Vizcaíno, Dirk Winkelhardt, and Veit Schwämmle*

Cite This: *J. Proteome Res.* 2024, 23, 418−429

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The inherent diversity of approaches in proteomics research has led to a wide range of software solutions for data analysis. These software solutions encompass multiple tools, each employing different algorithms for various tasks such as peptide-spectrum matching, protein inference, quantification, statistical analysis, and visualization. To enable an unbiased comparison of commonly used bottom-up label-free proteomics workflows, we introduce WOMBAT-P, a versatile platform designed for automated benchmarking and comparison. WOMBAT-P simplifies the processing of public data by utilizing the sample and data relationship format for proteomics (SDRF-Proteomics) as input. This feature streamlines the analysis of annotated local or public ProteomeXchange data sets, promoting efficient comparisons among diverse outputs. Through an evaluation using experimental ground truth data and a realistic biological data set, we uncover significant disparities and a limited overlap in the quantified proteins. WOMBAT-P not only enables rapid execution and seamless comparison of workflows but also provides valuable insights into the capabilities of different software solutions. These benchmarking metrics are a valuable resource for researchers in selecting the most suitable workflow for their specific data sets. The modular architecture of WOMBAT-P promotes extensibility and customization. The software is available at https://github.com/wombat-p/WOMBAT-Pipelines.



KEYWORDS: *workflow, data analysis, benchmarking, label-free proteomics, quality metrics*

INTRODUCTION

Computational workflows play a crucial role in data-intensive sciences such as mass spectrometry (MS)-based proteomics by providing a way to automate and streamline complex analysis processes and to make them easier to repeat and share with other researchers.[1] The search for an optimal data analysis solution is mostly data-dependent and cumbersome in many ways. In an optimal scenario, one would require extensive tests using ensembles of differently designed or parametrized workflows, all of them providing results in a comparable and standardized manner. We are aware that multiple search engines provide different results for the same data set, and how to integrate these results is an open question. In fact, on one hand, the combination of different workflows results in a boosted false identification rate of peptides and proteins; on the other hand, their intersection decreases the identification power. However, combining results from different tools remains a simple strategy to significantly improve the performance and reliability of the identification results for shotgun MS, maximizing the exploitation of the experimental MS spectra.[2]

Despite the significant developments made in recent years, there are still many challenges that must be addressed to enable more efficient, accurate, reproducible, and standardized analysis of proteomic data. This holds particularly for defining, building, executing, and benchmarking workflows. These challenges are not unique to proteomics but are exacerbated when compared to many other fields due to the diversity of experimental designs, operations on data, and file formats.

One major challenge is defining and identifying powerful algorithms and tools for proteomic data analysis. To achieve this, researchers require extensive knowledge of the current state of the art and software registries that provide an updated overview of the available tools such as bio.tools.[3] Additionally, benchmarks of software usage and popularity can help

Received: September 30, 2023
Revised: November 16, 2023
Accepted: November 22, 2023
Published: December 1, 2023

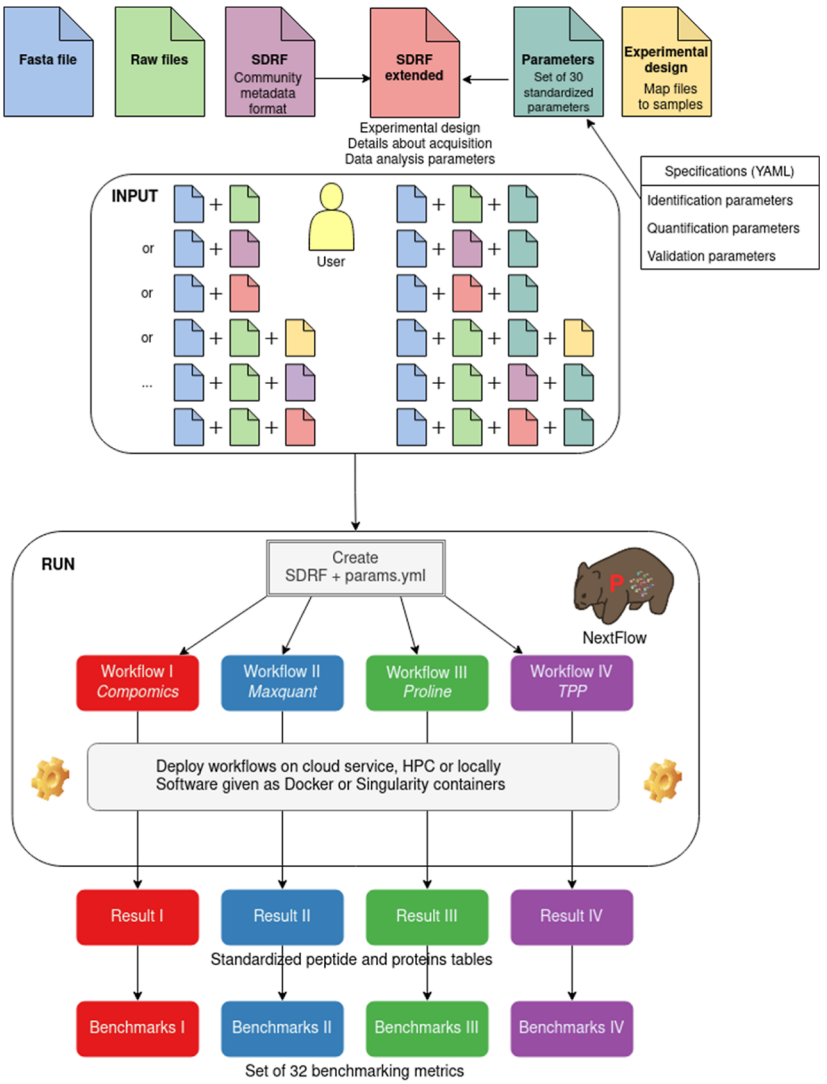


Figure 1. Scheme of the WOMBAT-P analysis workflows. They allow different types of input files for setting the workflow parameters and experimental design. This can be given either in the SDRF-Proteomics file format or as separate parameter files, which can also be used to overwrite the original settings.

researchers to identify the most appropriate tools for their specific research questions.[4]

Another challenge is constructing workflows from scratch or adding new software tools to an established pipeline. This can become problematic due to file format incompatibilities, inconsistent annotation, parameter definitions, and demands on the computational environment. However, an increasing amount of format converters and shims is starting to alleviate this issue. Moreover, accurate annotation of software tool input and output, e.g. via the EDAM ontology,[5] helps identifying suitable software combinations.[6,7] Interoperability issues can be solved using standardized file formats.

Running workflows relies on successful installations and execution settings to adapt to different computational environments. Workflow systems such as Galaxy, SnakeMake, common workflow language (CWL), and Nextflow can help address these issues by adapting the execution protocol to a wide range of local and cloud environments. These can utilize software containers such as Docker and Singularity[8,9] and standardized package management systems like Conda.[10] However, many proteomics tools are still not available via fully functional software containers or Conda packages.

Analyzing data of different origins can also pose a challenge as it requires the raw spectra and details about the study design and experimental protocol. Standardization of this information has been initiated recently via the SDRF-Proteomics format linking data files to samples and attributes from the data acquisition.[11,12] However, this format still lacks both in details about the data analysis protocol and availability in ProteomeXchange public repositories such as the PRIDE database.[13] Moreover, very few workflows can directly process this standardized information.

Workflow outputs like reports on peptide spectrum matches (PSMs), peptides, proteins, and differential abundance can come in a myriad of different formats and depend on different levels of interpretation, making a direct and objective comparison a major challenge. Efforts to compare the output of different data analysis pipelines have been carried out quite extensively (e.g., ref 14−16), but they are restricted to very few data sets. Nonetheless, these studies indicated large differences between the performance of various workflows, which shows the importance to benchmark different workflows and different workflow modules. Furthermore, few platforms provide benchmarking results over both different workflows and

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

different data sets without cumbersome adaptations to work with a particular data set.

While some efforts have been made to create central repositories for workflows, like Workflow Hub[17] and nf-core,[18] there are still relatively few workflows listed for proteomics there. This might be due to the relatively small size of the proteomics community as well as due to the heterogeneity and multitude of vendor-specific tools and workflows often used to characterize proteomic profiles.

To address these challenges, the WOMBAT-P platform has been developed to provide a comprehensive solution for defining, executing, and comparing ensembles of workflows. The platform captures both generic and user-defined specific benchmarks of performance, efficiency, and maintainability. These benchmarks are essential for evaluating workflow performance, its components, and hardware execution. They often use reference standard data sets and evaluate performance of quantitative preprocessing and statistical analysis such as a list of differentially abundant proteins, against which performance of other workflows can be measured. This ensures operation at an acceptable level and allows exploring the applicability of new software tools and their parameters over different data sets.

We demonstrate the WOMBAT-P platform with an ensemble of semantically equivalent and complete label-free quantification workflows for the analysis of bottom-up proteomics data. For this purpose, we used combinations of tools known from the recent literature to have a high degree of compatibility:

1. *Compomics* tools[19] + FlashLFQ[20] + MSqRob[21] (*Compomics* workflow).
2. MaxQuant[22] + NormalyzerDE[23] (MaxQuant workflow).
3. SearchGUI[24] + Proline[25] + PolySTest[26] (Proline workflow).
4. tools from the Trans-Proteomic Pipeline (TPP)[27] + ROTS[28] (TPP workflow).

METHODS

Workflow Implementation

WOMBAT-P bundles different workflows for the analysis of label-free proteomics data (Figure 1). It is built using Nextflow, a workflow language that allows running tasks across multiple computing infrastructures in a portable manner. We used an nf-core[18] template from 2021 to set up the main framework, and we used Docker and Singularity/Apptainer containers to make installation and results maximally reproducible. The Nextflow DSL2 implementation of the platform relies on one container image per process and allows describing each process in separate files, which makes it easier to maintain and update software dependencies. It also organizes the workflow steps as modules, facilitating their substitution or alternative tool combinations. As part of this study, we implemented four complete and mostly disjointed workflows based on existing tools. An overview of all the processes is available in Figure S1. We used the release version 0.9.2 in this study.

The first workflow (*Compomics* workflow) combines *Compomics* tools with FlashLFQ[20] and MSqRob.[21] The second workflow (*MaxQuant* workflow) is based on MaxQuant[22] and NormalyzerDE.[23] In the third workflow (*Proline* workflow), we used SearchGUI,[24] Proline,[25] and PolySTest.[26] Finally, we also included a workflow using tools from the TPP,[27] specifically PeptideProphet,[29] ProteinProphet, and StPeter[30] with Comet[31] for database search and ROTS[28] for statistical analysis (*TPP* workflow). Multiple modules were written in Python and R scripts to facilitate conversion and parametrization within the workflows. Therein, we used MSnbase[32] for normalization in the *Compomics* workflow and conversion scripts from https://github.com/jeffsocal/proteomic-id-tools in the TPP workflow.

Furthermore, we used wrProteo[33] for in-depth analysis of the investigated ground truth data set.

Input Options

WOMBAT-P allows a variety of input files, either based on SDRF-Proteomics annotation or via parameters given by a YAML file with the specifications of https://github.com/bigbio/proteomics-sample-metadata/blob/master/sdrf-proteomics/Data-analysis-metadata.adoc and the experimental design as a tab-delimited file (see Figure 1 for an overview of input options).

Workflow inputs are harmonized using a general set of parameters. Initialization and parametrization of the workflows are based on tools from the SDRF-pipelines[11] and the ThermoRawFileParser[34] for file conversion. We extended the definition of the SDRF-Proteomics data format to include a generalized set of 30 data analysis parameters (Table S1 and the specification of the YAML file), thus enabling the reproducibility of the data analysis via annotations with controlled vocabularies. When not provided, these parameters were set to their default values according to the specification file.

Output Options and Benchmarks

Intermediate and final files are stored in the results folder or a folder specified via the *outdir* parameter. In addition to the workflow-specific output, a standardized tabular format is provided at the peptide and protein levels (*stand_pep_quant_merged.csv* and *stand_prot_quant_merged.csv*, respectively).

For each of the workflows, WOMBAT-P calculates the same set of 32 benchmarking metrics for comparison between workflows and between different values of the data analysis parameters (Table S2).

Scripts for postprocessing are available at https://github.com/wombat-p/Utilities. We used the heatmap.2 function from the gplots R package for the hierarchical clustering. For the Reactome pathway enrichment analysis of the differentially abundant proteins resulting from each workflow,[35] the clusterProfiler Bioconductor R package[36] was used.

Benchmarking Data Sets

Raw data files were downloaded automatically from PRIDE by WOMBAT-P when SDRF-Proteomics files were used as the links to the files in the online repository are provided in these annotations. FASTA files for the database search were retrieved from UniProt (UniProtKB/SwissProt version Feb 3, 2023) and Sigma-Aldrich (https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/marketing/global/fasta-files/ups1-ups2-sequences.fasta).

The ground truth data set (PRIDE accession number PXD009815)[25] contains a yeast background and 48 human proteins (Universal Proteomics Data Set, UPS) spiked at 10 different concentrations of UPS proteins (10 amol, 50 amol, 100 amol, 250 amol, 500 amol, 1 fmol, 5 fmol, 10 fmol, 25 fmol, and 50 fmol).

For experimental data from a standard liquid chromatography (LC−MS) experiment, we used data from a study

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

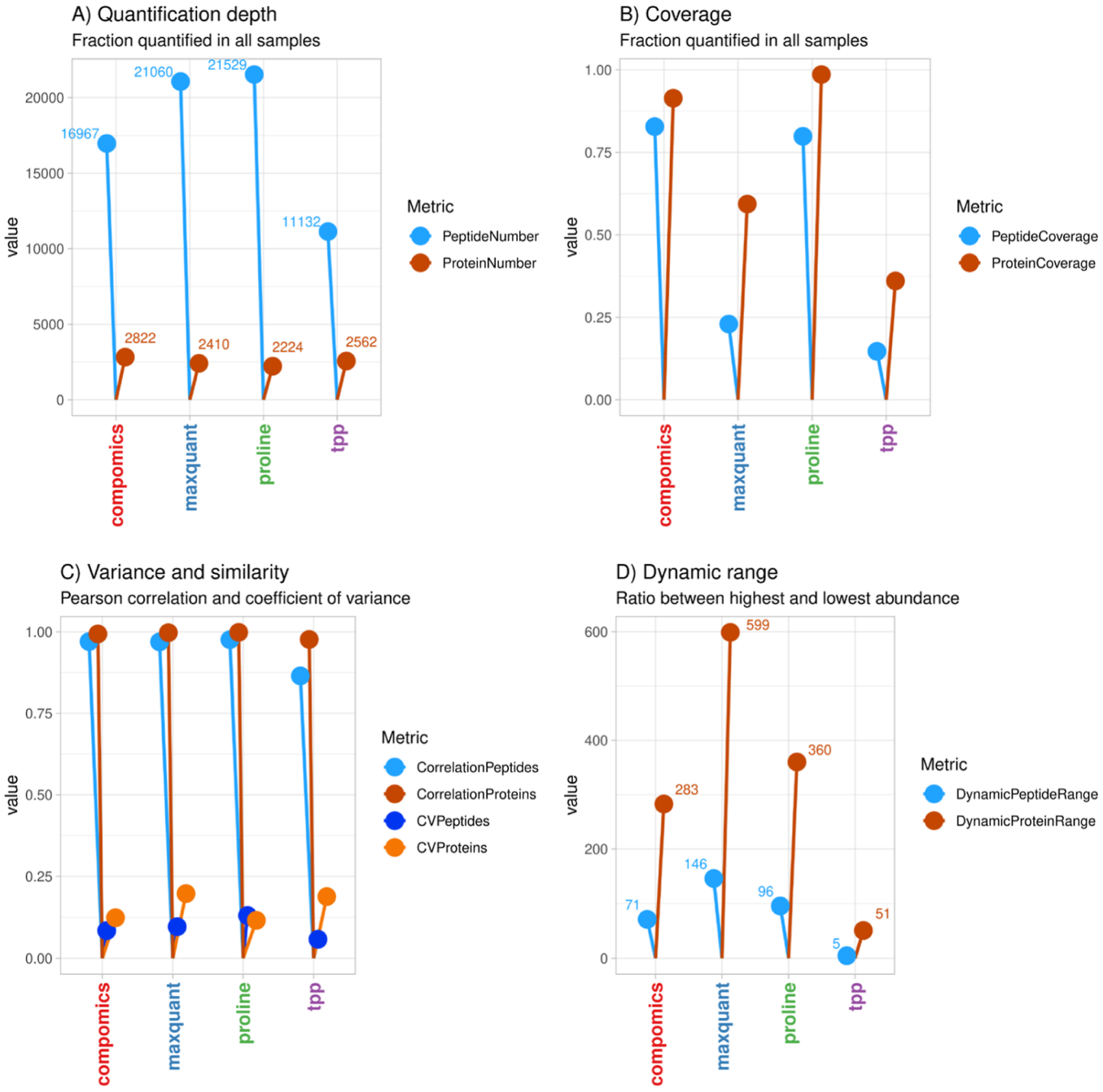


Figure 2. Summary of the benchmarking results for the ground truth data set.

comparing COVID-19-negative and -positive samples (PRIDE accession number PXD020394, from now on called COVID-19 data set).[37]

Workflow Registration

WOMBAT-P data provenance can be downloaded from the WorkflowHub as an RO-Crate (Research Object Crate[38]), see https://workflowhub.eu/workflows/444. The WOMBAT-P metadata RO-Crate contains information about the workflow, as well as its context. We used it to organize and share our workflow with other researchers in a standardized, interoperable, and reusable approach. The data provenance of WOMBAT-P was generated with WfExS-backend (https://github.com/inab/WfExS-backend), which is a high-level orchestrator to run scientific workflows reproducibly. It automates creating an RO-Crate by analyzing the structure and content of the computational workflow files.

Creating an RO-Crate of the WOMBAT-P workflow for instantiation or execution starts with the WOMBAT-P GitHub repository link. WfExS-backend analyzes the workflow repository, finds the workflow files, and extracts the metadata, including file names, formats, file sizes, script dependencies, and execution environments, needed to run the workflow. The extracted metadata is mapped to the corresponding RO-Crate metadata fields following the RO-Crate specification 1.1 (https://www.researchobject.org/ro-crate/1.1) to ensure that the metadata is correctly organized and represented. Then, a directory is generated using the extracted and mapped metadata with the necessary JSON-LD metadata files, including the *ro-crate-metadata.json* file, which contains the

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

comprehensive metadata for the RO-Crate. In addition, the directory includes the workflow files and associated data to ensure that all relevant files are included and can be accessed.

A full version of a generated Workflow Run RO-Crate from a workflow execution in singularity mode is available at https://doi.org/10.5281/zenodo.10091549. This RO-Crate contains snapshots of the workflow, inputs, containers, and results as payloads.

Availability

All workflows and documentation are available in GitHub at https://github.com/wombat-p/WOMBAT-Pipelines under an MIT license.

The main result files and the files needed for running the workflows are deposited at https://github.com/wombat-p/WOMBAT-P_Processed. While this article discusses the results from two particular data sets, the results from additional data sets are being added iteratively (five were available at the time of writing this article).

RESULTS

WOMBAT-P provides a platform for automated and scalable proteomics data analysis for bioinformaticians and experienced end-users, allowing a broad range of configurations to explore workflow performance thoroughly. These can be set using the 30 harmonized input parameters by specifying their values in either the SDRF file or the input parameter file. Parameters that are not provided are set to their default values (as documented in the WOMBAT-P GitHub main page).

The results, given in a standardized textual file format for peptide and protein quantification levels, can be assessed using 32 metrics to benchmark distinct features. These are given in the YAML format and thus can easily be processed for further in-depth evaluation. We ran the postprocessing using an R script, leading to the results described in what follows.

We provide detailed evaluations of the workflows and their results on the basis of two different data sets. WOMBAT version 0.9.2 was used for analyzing these data sets. The first data set is a ground truth data set with known information about the expected quantitative changes and thus serves to directly assess workflow performance. The second data set comes from a typical label-free proteomics experiment and thus should resemble the structure and properties of such.

Ground Truth Data Set

Comparison of different data analysis software and pipelines was performed on data from a ground-truth spike-in experiment.[25]

Regarding the overall number of quantified peptides and proteins, we observed that *TPP* reported a much lower peptide count than those of *MaxQuant* and *Compomics* (Figure 2A). We suspect that this is due to the prior filtering of peptides for protein false discovery rates (FDRs) in the *TPP* workflow. This also shows that the comparison at the same stage in the analysis is often hampered by even slightly different data treatment in the workflows. Despite the differences in peptide quantifications, all workflows showed similar numbers of proteins (Figure 2A). However, this means that the proportionality of the reported quantified peptides and proteins differs between the different tested pipelines. This discrepancy implies that the underlying tools employ algorithms (such as protein inference and protein-level FDR control), leading to a different number of protein groups for a given number of peptides.

Proline and FlashLFQ from the *Compomics* workflow reported higher coverage of peptides and proteins across samples compared to that of *TPP*, which lacked a match-between-runs option (Figure 2B). A lower coverage of quantified peptides across samples was observed in *MaxQuant*, and even lower in *TPP*. The lower number in *TPP* is likely due to the inability to use information from other MS runs to improve the overall number of quantified peptides and proteins (Figure 2B).

Higher coverage decreased the variation of protein abundance values between replicates, and we consistently observed high correlations for all workflows with the exception of *TPP*, which provided slightly lower performance (Figure 2C).

Finally, we evaluated the dynamic range, i.e., the ratio between the highest and the lowest reported quantitative values for the reported peptides and proteins, to reproduce actual changes in protein and peptide abundance within the mass spectrometer's sensitivity. We found that *MaxQuant* had an approximately 10-fold higher range of protein abundance changes of 599 when compared to that in *TPP* with 51, which is considerably different from the variability observed in the other workflows (Figure 2D). Here, StPeter from *TPP* uses a method of quantification different from that used by the other tools. It is noteworthy to mention that a higher dynamic range does not necessarily reflect a more accurate output as distinct summarization of PSMs and peptides can lead to different deviations from the linear response.

While generic benchmarking metrics such as the number of quantified peptides and proteins offer valuable insights into workflow performance, it is equally important to assess the overlap and similarity between the obtained results. We then evaluated the overlap in quantified proteins across all the samples and workflows (Figure S2), revealing distinct patterns both across and within the four workflows. The *Compomics* and *Proline* workflows exhibited a higher overlap among the different samples.

In terms of quantified peptides, we observed that their similarity with respect to their relative abundances was generally higher within each workflow (Figure S3), likely due to variations in the methods used to quantify peptide and protein intensities. This trend was particularly noticeable in the *TPP* workflow, where quantification was based on StPeter, an algorithm that incorporates spectral counting. Furthermore, the workflows successfully grouped the samples with the same UPS protein concentrations in most instances.

The samples of the experiment consist of 48 human proteins spiked at levels of the yeast proteome. Thus, only all proteins annotated as this species (human in this data set) are expected to vary between the samples, while the proteins of the other species (yeast in this data set) are expected to be always detected at a constant level.

We run the workflows with median normalization, i.e., without using a specialized normalization method that adapts to the rather particular experimental design. Given the large concentration changes of the spike-in UPS proteins, this design led to an apparent underexpression of the background proteins in the conditions where the UPS proteins are highly abundant, thus providing misleading quantitative changes of the background proteins. A specialized analysis and extensive interpretations using wrProteo for this spike-in ground truth data including use of appropriate normalization are available in the Supporting Information File.

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

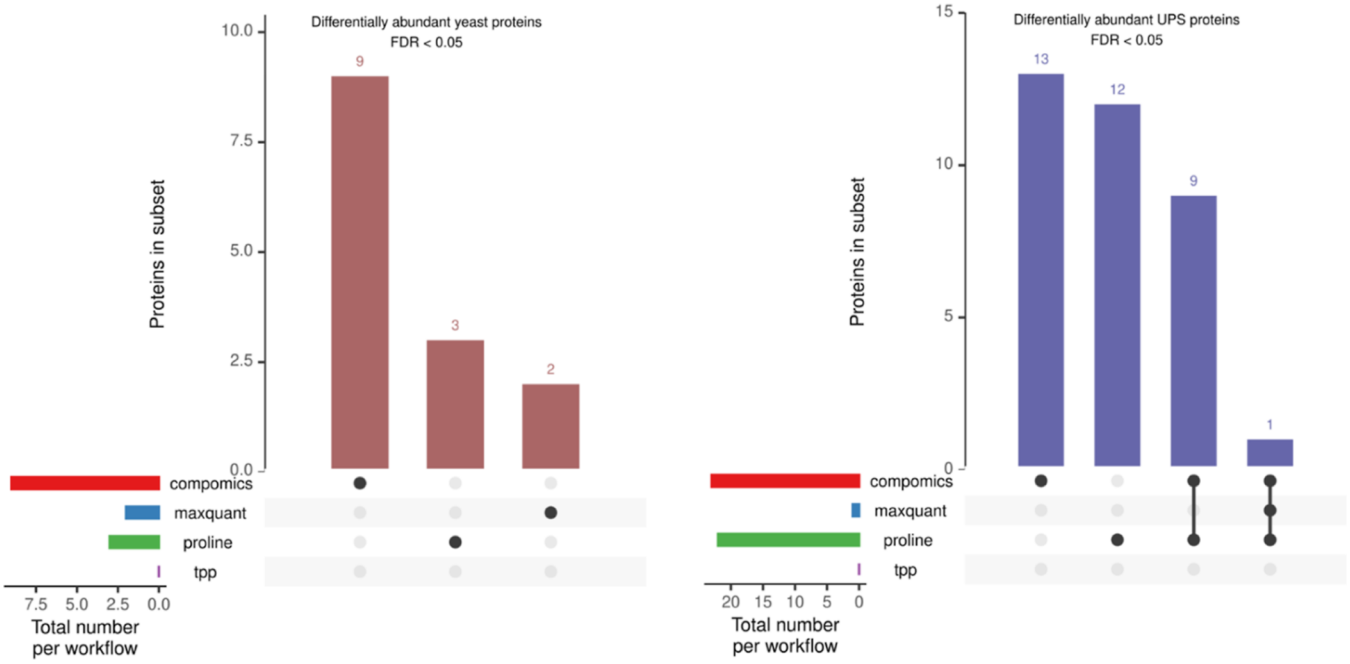


Figure 3. Comparison of the proteins found to be differentially abundant between the samples with UPS amounts of 10 and 500 amol and a background of equally abundant yeast proteins. Differential regulation was determined by the respective workflow components for statistical testing, which uses different statistical approaches.

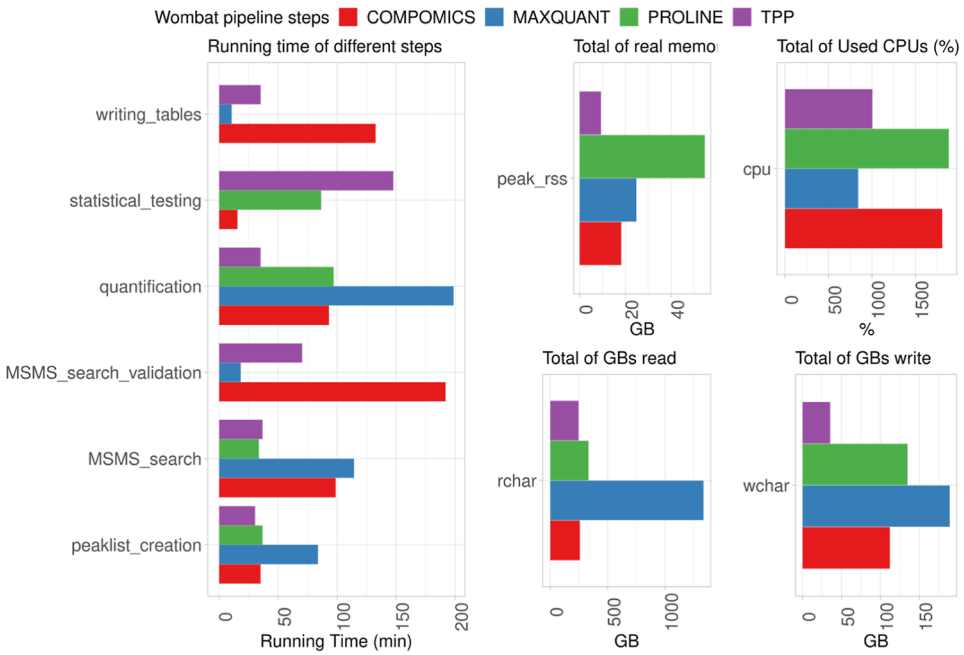


Figure 4. Comparison of the computational resources used for the different workflows. Given that *MaxQuant* bundles multiple operations, we only report running times, which were available in the *MaxQuant* output. rchar: amount of the data being read; wchar: amount of written data; peak_rss: peak amount of RAM; CPU: average number of used CPU threads in percentage.

However, in a regular experiment, it is not known in advance as to which proteins are expected to be constant in abundance. Therefore, we did not apply such data transformations in order to provide an objective view of the workflow performance. This also meant that the "official" ground truth of finding 48 differentially abundant UPS proteins became mixed with differentially abundant yeast proteins. While assuming that this should not affect the outcome too drastically, we find surprisingly different results when comparing the output from the four WOMBAT-P workflows.

We checked how well the workflows detected the UPS proteins as differentially abundant in a challenging case of their low concentrations being 10 versus 500 amol. For that, we assessed the number of human proteins observed as variants (providing the sensitivity) and the percentage of wrongly detected yeast variants (providing the specificity). When examining the outputs of the workflows in terms of

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

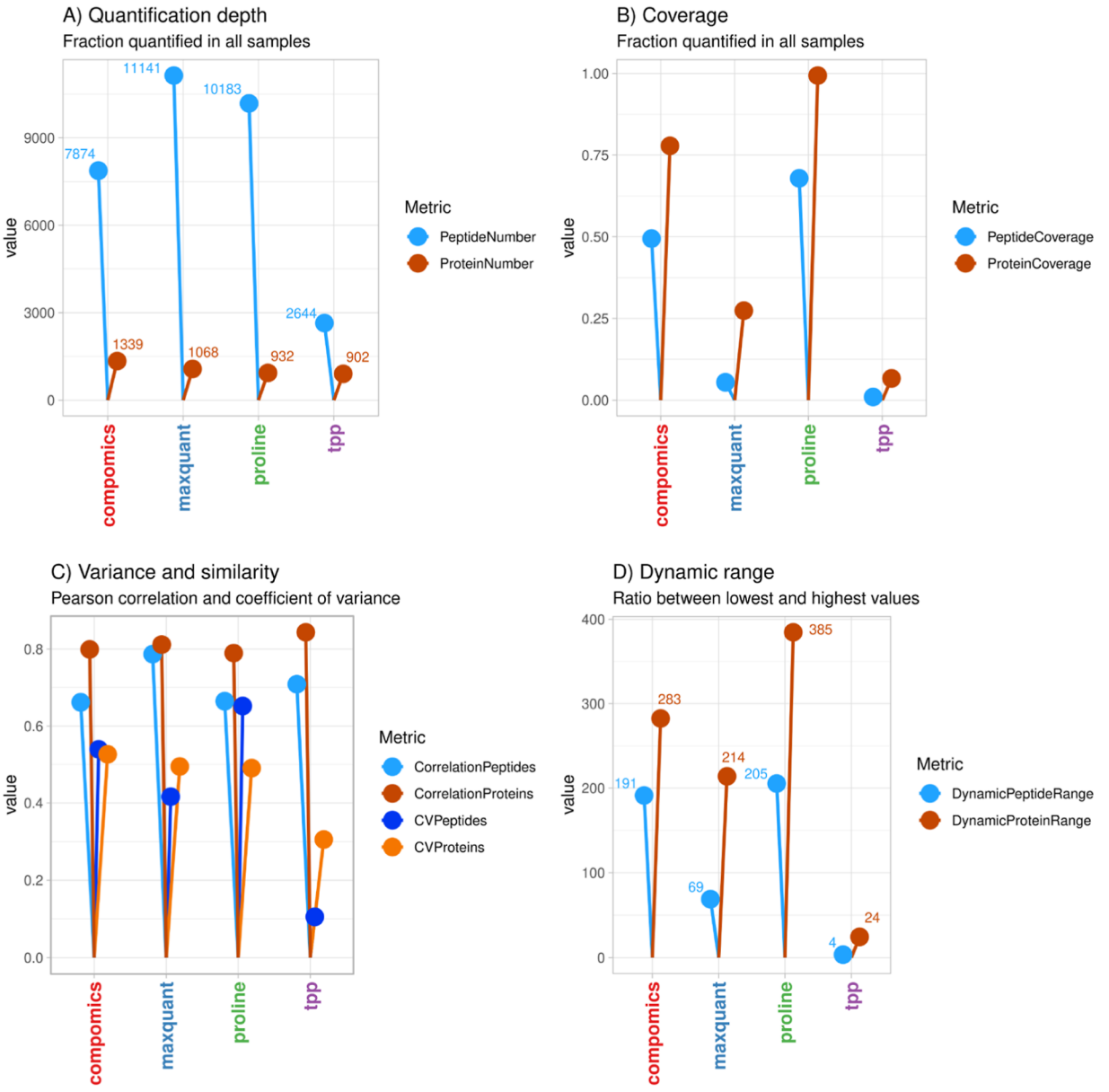


Figure 5. Summary of the benchmarking results for biological data set.

differentially abundant proteins (FDR < 0.05), there was poor agreement among all workflows (Figure 3). The *Proline* and *Compomics* workflows showed the highest number of UPS proteins correctly detected as differentially abundant with a total of 37 and 35 proteins, respectively. Notably, 14 of the 37 UPS proteins identified by *Proline* as significantly different were found only by this particular workflow. Similarly, the *Compomics* workflow reported 12 UPS proteins uniquely found in this workflow. When comparing with the number of differentially abundant yeast proteins representing false-positives, their numbers were lower, and there was no overlap between the workflows. Notably, all eight proteins found in at least two workflows were UPS proteins. For a more systematic exploration of the ground truth, we refer to Supporting Information File 1.

We also measured the usage of available computational resources with different metrics (Figure 4). These metrics were extracted from a trace report that was computed by nextflow. It contains information about each executed process in the pipeline. The results varied widely across different tasks, which could be attributed to different implementations and differently assigned subtasks of the major data transformations.

On the whole, we noticed that *Proline* and *Compomics* used CPUs more efficiently and required more memory, while the *MaxQuant* workflow performed more file read/write operations. In summary, when utilizing the ground truth data set with UPS proteins spiked into a yeast background, significant differences were observed among the outputs of the various workflows.

Performance on Biological Data from COVID-19 Study

Ground truth data sets can be limited in accurately replicating biological samples. Therefore, we decided to also assess the performance of WOMBAT-P using a recent data set that compared COVID-19-positive and -negative samples.[37] The utilization of this data set led to slightly different results, showing that the workflow performance can depend strongly on each particular data set it processes.

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

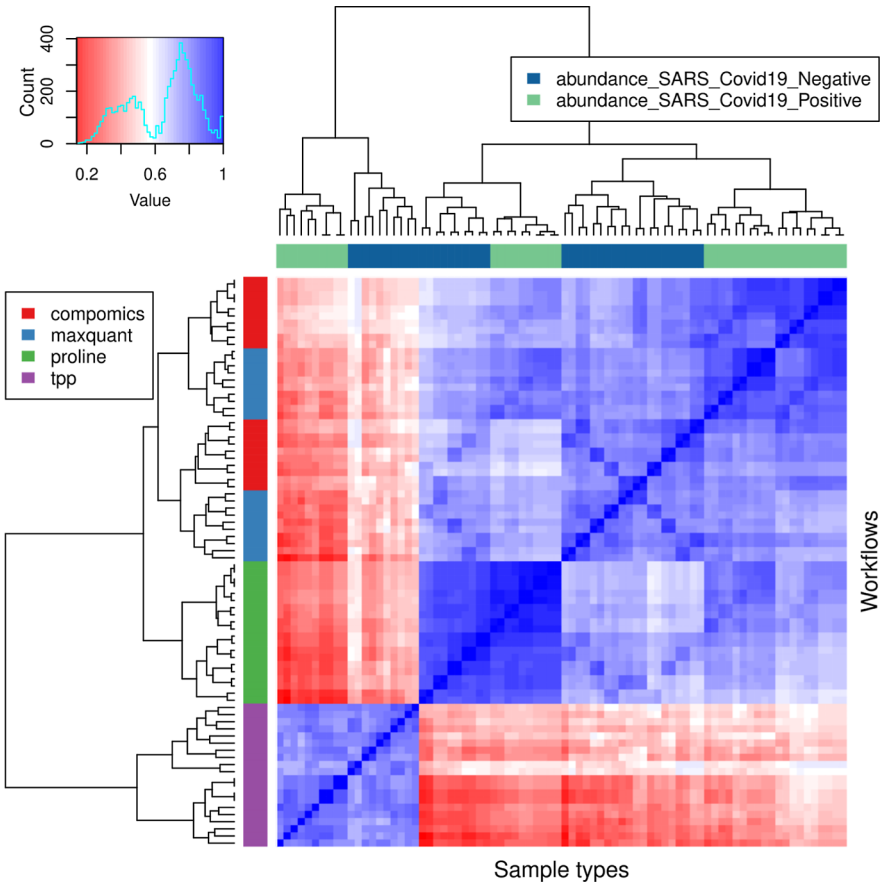


Figure 6. Similarity of workflow results using the COVID-19 data set. Pearson correlation values were calculated to compare the similarity on a quantitative level, and then, the correlations were arranged using hierarchical clustering. In contrast to the benchmarking metrics in Figures 2 and 5, here, the Pearson correlation values were calculated from the $\log_2$-transformed values.

When comparing the different benchmarks (Figure 5), the *Compomics* and *TPP* workflows consistently exhibited lower peptide quantification numbers, whereas the *Compomics* workflow demonstrated up to 400 more quantified proteins compared to that by the other workflows. When evaluating the variance and correlation levels within the replicates of the same sample type, the correlations were lower compared to that with the ground truth data. This disparity can be attributed to the increased biological variability among the samples (from controls and patients) and to the lower number of quantified proteins. Notably, the *TPP* workflow exhibited the lowest coefficient of variance and the highest protein abundance value correlations of nonlog-transformed values. It is worth mentioning that we observed significant differences in the dynamic ranges of protein and peptide quantifications, similar to what was observed in the ground truth data set. These differences could have influenced the high correlation values and low coefficient of variance for the *TPP* workflow.

Run times showed a diverse picture, similarl to what was observed for the UPS data set (Figure S4). Additionally, computer resources used for the processing of this data set exhibited very close relative values, compared to those used for the UPS data set.

When comparing the overlap of the quantified proteins within and across the workflows (Figure S5), it could be confirmed that there was a much higher similarity in the proteomes measured among the samples analyzed with the same workflow. This cannot be merely attributed to the applied match-between-runs option, given that the *TPP* workflow separates similarly. While it is well attested in the literature[39−41] that different search engines identify non-identical sets of peptides from the same data, downstream components, such as for PSM validation[42] or protein inference,[43] may also be of considerable influence.

The quantitative comparison of the workflow results (Figure 6) provided more insights into the workflow performance. For the correlations within the samples of a given workflow, the *Proline* workflow performed with the highest similarity among all the samples. Notably, the results from the *Compomics* and *MaxQuant* workflows were sufficiently similar to separate the two sample types between them.

We furthermore compared the protein groups found to be differentially abundant (Figure S6). While more than 100 proteins were detected in at least 3 of the 4 workflows, they disagreed to a high degree, leading to more than 500 proteins being uniquely found to be significantly changing between the COVID-19-positive and -negative samples. However, further analysis via Reactome pathway enrichment led to considerable agreement between the results of the *Compomics*, *Maxquant*, and *Proline* workflows for the most enriched pathways (Figure S7).

In summary, we found both similarities and differences when comparing workflow performances. The noise levels, as expected, were higher for the COVID-19 data set and showed higher similarity between the different workflows (Figures 2C and 5C).

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

DISCUSSION

This study introduces WOMBAT-P as an innovative solution for addressing the challenges of large-scale proteomic data analysis. It relies on the importance of automated data quality control and validation in scaling up to analyze a large number of files. The scalability of WOMBAT-P using high-performance computing (HPC) environments and its utilization of software containers enable reproducible analyses, making it a valuable tool for both public and in-house label-free data analysis on a large scale. WOMBAT-P provides a directly applicable, command-line-based platform for the analysis of in-house data and for the reanalysis of public data sets. As the workflows are written in NextFlow, it can furthermore be embedded in cloud and HPC environments.

One notable feature of WOMBAT-P is its ability to create SDRF-Proteomics file templates and produce harmonized outputs, facilitating easy benchmarking and a comparison of results. This comprehensive approach enhances the robustness and reliability of proteomic data analysis.

We found different performances of the workflows when testing them on two selected data sets. The comparison of the quantification depth (ground truth data set) revealed discrepancies in the reported number of protein groups, when relatively compared to the number of peptides. This could be attributed to various factors, primarily stemming from the intricacies of the employed protein inference algorithms and differences in FDR control at the protein identification level. To delve deeper into this issue, it would be valuable to examine the occurrence of one-hit wonders peptides (that uniquely identify a protein) in each pipeline and also to perform tandem mass spectrometry searches using entrapment databases[44,45] in order to better understand the nuances of the underlying protein identification algorithms. The quantitative performance comparison provided further details. The *Compomics* and *Proline* workflows showed relatively higher differentially abundant protein numbers, likely due to a more effective match-between-runs feature. This finding of rather different performances suggests the absence of a universally optimal solution for proteomic data analysis, which might limit the routine use of proteomics in clinical settings. It underscores the significance of data-driven workflow analysis using benchmarking metrics to assess and identify the best-performing solutions for proteomic data analysis. However, the biological interpretation of the results of the COVID-19 data set also showed that despite providing rather different results, the workflow outputs still describe very similar biological pathways, and thus, these differences might not matter too much for systems in biology studies. However, this observation most likely is not generalizable to other kinds of studies, such as, for instance, biomarker discoveries and quantitative profiling of post-translational modifications.

We are aware that benchmarking of software can be complex due to the number of options available in terms of parameters, the different characteristics of the benchmarking data sets, and the need for expertise in all the tools used in the benchmarking. WOMBAT-P allows users to run and compare different configurations and thus explore alternative and more optimized setup.

By providing a large number of 32 benchmarking metrics, the user can tailor their assessment to their specific requirements, such as the highest quantification depth, the highest confidence in identifying differentially abundant proteins, or the highest overlap of quantified proteins within all samples. On purpose, given the often very different needs, we cannot provide an optimal solution. Therefore, the user must have a clear idea of what they consider to be optimal performance depending on their needs and requirements.

Finding an optimal solution for a data analysis depends not only on selecting the software and algorithms but also on fine-tuning the parameter settings. The scalability of WOMBAT-P to larger computational resources allows running multiple instances to compare them on the basis of the provided benchmarking metrics. Thus, the user can perform ensembles of workflow executions to study both the performance and robustness of the results of a particular data set.

In addition to finding the workflow solution that performs best for a given set of criteria, by allowing to run different workflows on the same data set, WOMBAT-P facilitates combining the results from multiple workflows and thus promotes, e.g., investigations of the complementarity of different algorithmic and statistical approaches. Such combinations can be done at different levels during the analysis and thus increase the complexity considerably. Moreover, combining different results will require further processing, such as careful correction for multiple testing.

The modularized architecture of WOMBAT-P enables the incorporation of new processing algorithms and steps and workflows, including new advancements such as tools that apply deep learning and further downstream software for biological interpretation. This flexibility enhances the capabilities of WOMBAT-P and expands its potential for further developments in proteomic analyses.

We view WOMBAT-P as a continuously evolving platform with ongoing development and open development, with new modules and updates anticipated in the near future, and therefore, we invite contributions from the proteomics bioinformatics community. Achieving the exchangeability of certain operations like the downstream statistical tests across all four workflows will make WOMBAT-P a more versatile and flexible tool for proteomic data analysis.

In conclusion, this study highlights the significance and relevance of WOMBAT-P for proteomic data analysis. By providing a comprehensive tool that enhances accuracy, scalability, and comparability in the analysis of large-scale proteomic analyses, WOMBAT-P addresses critical challenges and contributes to advancing our understanding of complex biological systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.3c00636.

Flowchart of all processes run by WOMBAT-P, overlap of protein identifications between all ground truth data samples analyzed by the different workflows, similarity of protein quantitative profiles between different samples of the ground truth dataset, running times for the COVID-19 data set, overlap of identified proteins in the different samples and workflows for the COVID-19 data set, differentially abundant proteins for COVID 19 dataset for an FDR threshold of 5%, and most enriched Reactome pathways for proteins found to be differentially abundant between COVID-19-positive and -negative samples. (PDF)

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

Definitions and categories of data analysis parameters in WOMBAT-P (XLSX)

Definitions and categories of benchmarking metrics in WOMBAT-P (XLSX)

Detailed evaluation of the UPS data set with wrProteo (PDF)

AUTHOR INFORMATION

Corresponding Author

Veit Schwämmle − *Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark;* orcid.org/0000-0002-9708-6722; Email: veits@bmb.sdu.dk

Authors

David Bouyssié − *Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, Université Toulouse III—Paul Sabatier (UT3), 31062 Toulouse, France; Proteomics French Infrastructure, ProFI, FR 2048 Toulouse, France;* orcid.org/0000-0002-0847-4759

Pınar Altıner − *Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, Université Toulouse III—Paul Sabatier (UT3), 31062 Toulouse, France*

Salvador Capella-Gutierrez − *Life Sciences Department, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain*

José M. Fernández − *Life Sciences Department, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain;* orcid.org/0000-0002-4806-5140

Yanick Paco Hagemeijer − *Department of Analytical Biochemistry, University of Groningen, Groningen Research Institute of Pharmacy, 9712 CP Groningen, The Netherlands; European Research Institute for the Biology of Ageing, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands;* orcid.org/0000-0001-6036-3741

Peter Horvatovich − *Department of Analytical Biochemistry, University of Groningen, Groningen Research Institute of Pharmacy, 9712 CP Groningen, The Netherlands;* orcid.org/0000-0003-2218-1140

Martin Hubálek − *Institute of Organic Chemistry and Biochemistry, CAS, 160 00 Prague, Czech Republic;* orcid.org/0000-0003-0247-7956

Fredrik Levander − *National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Immunotechnology, Lund University, 22100 Lund, Sweden;* orcid.org/0000-0002-0710-9792

Pierluigi Mauri − *Institute for Biomedical Technologies (ITB), Department of Biomedical Sciences, National Research Council (CNR), 20054 Milan, Italy*

Magnus Palmblad − *Leiden University Medical Center, 2300 RC Leiden, The Netherlands;* orcid.org/0000-0002-5865-8994

Wolfgang Raffelsberger − *Wolfgang Raffelsberger: Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université de Strasbourg, CNRS UMR7104, INSERM U1258, Illkirch, 67404 Illkirch, France*

Laura Rodríguez-Navas − *Life Sciences Department, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain*

Dario Di Silvestre − *Institute for Biomedical Technologies (ITB), Department of Biomedical Sciences, National Research Council (CNR), 20054 Milan, Italy*

Balázs Tibor Kunkli − *Balázs Tibor Kunkli: Department of Biochemistry and Molecular Biology, University of Debrecen, 4032 Debrecen, Hungary;* orcid.org/0000-0003-1266-2792

Julian Uszkoreit − *Medical Faculty, Medical Bioinformatics, Center for Protein Diagnostics (ProDi), Medical Proteome Analysis, and Medical Faculty, Medizinisches Proteom-Center, Ruhr University Bochum, 44801 Bochum, Germany;* orcid.org/0000-0001-7522-4007

Yves Vandenbrouck − *Proteomics French Infrastructure, ProFI, FR 2048 Toulouse, France; CEA, Fundamental Research Division, Proteomics French Infrastructure, 91191 Gif-sur-Yvette, France;* orcid.org/0000-0002-1292-373X

Juan Antonio Vizcaíno − *European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI), Cambridge CB10 1SD, U.K.;* orcid.org/0000-0002-3905-4335

Dirk Winkelhardt − *Medical Faculty, Medizinisches Proteom-Center, Ruhr University Bochum, 44801 Bochum, Germany;* orcid.org/0000-0001-8770-2221

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.3c00636

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by ELIXIR (https://elixir-europe.org/), the European research infrastructure for life-science data. The work was also funded in part by a grant from the French Ministry of Research with the "Investissement d'Avenir Infrastructures Nationales en Biologie et Santé" program (ProFI, Proteomics French Infrastructure project, ANR-10-INBS-08). J.A.V. would also like to acknowledge the funding from BBSRC [grant number BB/T019670/1].

REFERENCES

(1) Harjes, J.; Link, A.; Weibulat, T.; Triebel, D.; Rambold, G. FAIR digital objects in environmental and life sciences should comprise workflow operation design data and method information for repeatability of study setups and reproducibility of results. *Database* **2020**, *2020*, baaa059.

(2) Zhao, P.; Zhong, J.; Liu, W.; Zhao, J.; Zhang, G. Protein-Level Integration Strategy of Multiengine MS Spectra Search Results for Higher Confidence and Sequence Coverage. *J. Proteome Res.* **2017**, *16*, 4446−4454.

(3) Ison, J.; Ienasescu, H.; Chmura, P.; Rydza, E.; Ménager, H.; Kalaš, M.; Schwämmle, V.; Grüning, B.; Beard, N.; Lopez, R.; et al. The bio.tools registry of software tools and data resources for the life sciences. *Genome Biol.* **2019**, *20*, 164.

(4) Capella-Gutierrez, S.; Iglesia, D. d. l.; Haas, J.; Lourenco, A.; Fernández, J. M.; Repchevsky, D.; Dessimoz, C.; Schwede, T.; Notredame, C.; Gelpi, J. L.; et al. Lessons learned: Recommendations for establishing critical periodic scientific benchmarking. *bioRxiv* **2017**, 181677.

(5) Ison, J.; Kalaš, M.; Jonassen, I.; Bolser, D.; Uludag, M.; McWilliam, H.; Malone, J.; Lopez, R.; Pettifer, S.; Rice, P. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics* **2013**, *29*, 1325−1332.

(6) Palmblad, M.; Lamprecht, A.-L.; Ison, J.; Schwämmle, V. Automated workflow composition in mass spectrometry-based proteomics. *Bioinformatics* **2019**, *35*, 656−664.
(7) Kasalica, V.; Schwämmle, V.; Palmblad, M.; Ison, J.; Lamprecht, A.-L. APE in the Wild: Automated Exploration of Proteomics Workflows in the bio.tools Registry. *J. Proteome Res.* **2021**, *20*, 2157−2165.
(8) da Veiga Leprevost, F.; Grüning, B. A.; Alves Aflitos, S.; Röst, H. L.; Uszkoreit, J.; Barsnes, H.; Vaudel, M.; Moreno, P.; Gatto, L.; Weber, J.; et al. BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics* **2017**, *33*, 2580−2582.
(9) Bai, J.; Bandla, C.; Guo, J.; Vera Alvarez, R.; Bai, M.; Vizcaíno, J. A.; Moreno, P.; Grüning, B.; Sallou, O.; Perez-Riverol, Y. BioContainers Registry: Searching Bioinformatics and Proteomics Tools, Packages, and Containers. *J. Proteome Res.* **2021**, *20*, 2056−2061.
(10) Grüning, B.; Dale, R.; Sjödin, A.; Chapman, B. A.; Rowe, J.; Tomkins-Tinch, C. H.; Valieris, R.; Köster, J. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **2018**, *15*, 475−476.
(11) Perez-Riverol, Y.; European Bioinformatics Community for Mass Spectrometry. Toward a Sample Metadata Standard in Public Proteomics Repositories. *J. Proteome Res.* **2020**, *19*, 3906−3909.
(12) Dai, C.; Füllgrabe, A.; Pfeuffer, J.; Solovyeva, E. M.; Deng, J.; Moreno, P.; Kamatchinathan, S.; Kundu, D. J.; George, N.; Fexova, S.; et al. A proteomics sample metadata representation for multiomics integration and big data analysis. *Nat. Commun.* **2021**, *12*, 5854.
(13) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543−D552.
(14) Lietzén, N.; Natri, L.; Nevalainen, O. S.; Salmi, J.; Nyman, T. A. Compid: a new software tool to integrate and compare MS/MS based protein identification results from Mascot and Paragon. *J. Proteome Res.* **2010**, *9*, 6795−6800.
(15) Hoekman, B.; Breitling, R.; Suits, F.; Bischoff, R.; Horvatovich, P. msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies. *Mol. Cell. Proteomics* **2012**, *11*, M111.015974.
(16) Locard-Paulet, M.; Bouyssié, D.; Froment, C.; Burlet-Schiltz, O.; Jensen, L. J. Comparing 22 Popular Phosphoproteomics Pipelines for Peptide Identification and Site Localization. *J. Proteome Res.* **2020**, *19*, 1338−1345.
(17) Goble, C.; et al. *Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory*, 2021, Preprint at..
(18) Ewels, P. A.; Peltzer, A.; Fillinger, S.; Patel, H.; Alneberg, J.; Wilm, A.; Garcia, M. U.; Di Tommaso, P.; Nahnsen, S. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **2020**, *38*, 276−278.
(19) Barsnes, H.; Vaudel, M.; Colaert, N.; Helsens, K.; Sickmann, A.; Berven, F. S.; Martens, L. compomics-utilities: an open-source Java library for computational proteomics. *BMC Bioinf.* **2011**, *12*, 70.
(20) Millikin, R. J.; Shortreed, M. R.; Scalf, M.; Smith, L. M. Fast, Free, and Flexible Peptide and Protein Quantification with FlashLFQ. *Methods Mol. Biol.* **2023**, *2426*, 303−313.
(21) Goeminne, L. J. E.; Gevaert, K.; Clement, L. Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. *J. Proteonomics* **2018**, *171*, 23−36.
(22) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301−2319.
(23) Willforss, J.; Chawade, A.; Levander, F. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J. Proteome Res.* **2019**, *18*, 732−740.
(24) Barsnes, H.; Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J. Proteome Res.* **2018**, *17*, 2552−2555.
(25) Bouyssié, D.; Hesse, A. M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez de Peredo, A.; Couté, Y.; Dupierris, V.; Burel, A.; et al. Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020**, *36*, 3148−3155.
(26) Schwämmle, V.; Hagensen, C. E.; Rogowska-Wrzesinska, A.; Jensen, O. N. PolySTest: Robust Statistical Testing of Proteomics Data with Missing Values Improves Detection of Biologically Relevant Features. *Mol. Cell. Proteomics* **2020**, *19*, 1396−1408.
(27) Deutsch, E. W.; Mendoza, L.; Shteynberg, D. D.; Hoopmann, M. R.; Sun, Z.; Eng, J. K.; Moritz, R. L. Trans-Proteomic Pipeline: Robust Mass Spectrometry-Based Proteomics Data Analysis Suite. *J. Proteome Res.* **2023**, *22*, 615−624.
(28) Suomi, T.; Seyednasrollah, F.; Jaakkola, M. K.; Faux, T.; Elo, L. L. ROTS: An R package for reproducibility-optimized statistical testing. *PLoS Comput. Biol.* **2017**, *13*, No. e1005562.
(29) Ma, K.; Vitek, O.; Nesvizhskii, A. I. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinf.* **2012**, *13*, S1.
(30) Hoopmann, M. R.; Winget, J. M.; Mendoza, L.; Moritz, R. L. StPeter: Seamless Label-Free Quantification with the Trans-Proteomic Pipeline. *J. Proteome Res.* **2018**, *17*, 1314−1320.
(31) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13*, 22−24.
(32) Gatto, L.; Gibb, S.; Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J. Proteome Res.* **2021**, *20*, 1063−1069.
(33) Raffelsberger, W. *wrProteo: Proteomics Data Analysis Functions*, 2023. https://CRAN.R-project.org/package=wrProteo.
(34) Hulstaert, N.; Shofstahl, J.; Sachsenberg, T.; Walzer, M.; Barsnes, H.; Martens, L.; Perez-Riverol, Y. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J. Proteome Res.* **2020**, *19*, 537−542.
(35) Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **2022**, *50*, D687−D692.
(36) Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141.
(37) Rivera, B.; Leyva, A.; Portela, M. M.; Moratorio, G.; Moreno, P.; Durán, R.; Lima, A. Quantitative proteomic dataset from oro- and naso-pharyngeal swabs used for COVID-19 diagnosis: Detection of viral proteins and host's biological processes altered by the infection. *Data Brief* **2020**, *32*, 106121.
(38) Soiland-Reyes, S.; Sefton, P.; Crosas, M.; Castro, L. J.; Coppens, F.; Fernández, J. M.; Garijo, D.; Grüning, B.; La Rosa, M.; Leo, S.; et al. Packaging research artefacts with RO-Crate. *Data Sci.* **2022**, *5*, 97−138.
(39) Mohammed, Y.; Palmblad, M. Visualizing and comparing results of different peptide identification methods. *Briefings Bioinf.* **2018**, *19*, 210−218.
(40) De La Toba, E. A.; Anapindi, K. D. B.; Sweedler, J. V. Assessment and Comparison of Database Search Engines for Peptidomic Applications. *J. Proteome Res.* **2023**, *22*, 3123−3134.
(41) Yuan, Z.-F.; Lin, S.; Molden, R. C.; Garcia, B. A. Evaluation of proteomic search engines for the analysis of histone modifications. *J. Proteome Res.* **2014**, *13*, 4470−4478.
(42) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **2011**, *10*, M111.007690.

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429

(43) Audain, E.; Uszkoreit, J.; Sachsenberg, T.; Pfeuffer, J.; Liang, X.; Hermjakob, H.; Sanchez, A.; Eisenacher, M.; Reinert, K.; Tabb, D. L.; et al. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics. *J. Proteomics* **2017**, *150*, 170−182.
(44) Granholm, V.; Noble, W. S.; Käll, L. On using samples of known protein content to assess the statistical calibration of scores assigned to peptide-spectrum matches in shotgun proteomics. *J. Proteome Res.* **2011**, *10*, 2671−2678.
(45) Ebadi, A.; Freestone, J.; Noble, W. S.; Keich, U. Bridging the False Discovery Gap. *J. Proteome Res.* **2023**, *22*, 2172−2178.

https://doi.org/10.1021/acs.jproteome.3c00636
J. Proteome Res. 2024, 23, 418−429