nature portfolio

Peer Review File

OmicScope unravels systems-level insights from quantitative proteomics data

Reviewer #1 (Remarks to the Author):

Synopsis:
OmicScope does facilitate proteomics analyses, making some complex methods simpler. There are no new implementations, the full field of similar tools has not been adequately addressed, and the demonstrations are lacking compelling comparisons to what was previously published (e.g was something confirmed or newly discovered). With added clarity and revisions this reviewer might accept publishing.

Summary:
OmicScope appears to be a well formulated python implementation for post analysis of proteomic data with both a programming code base and web implementation. Of particular interest is the multi-omics and longitudinal analyses, along with the demonstrated analysis of available open access data. However, the authors fail to demonstrate a longitudinal analysis, and the multi-dataset demonstration did not highlight new insights, something that would be compelling for the readers.

In addition, the authors fail to address in the introduction, the myriad other software implementations in python and other languages, or define distinguishing characteristics of their implementation from what is already available among a congested field of proteomic analysis tools. Instead leaving a partial survey of similar tools for the discussion. OmicScope applies well worn analyses to quantified proteomics data, however, it offers no quality assessment, manual filtering capabilities, or data treatments, therefore their claim as a "comprehensive solution" seems unwarranted. The authors are advised to narrow their claims, or demonstrate more complete functionality.

Specifics:
Authors claim "OmicScope emerges as a comprehensive solution for quantitative proteomics data analysis"[line 18], claiming inputs from a handful of platforms, however, they fail to address widely utilized proteomics platforms such as ProteomeDiscoverer, FragPipe and OpenMS and instead offer a only a generalized input function. Authors should address these platforms, and offer some path to availability to claim "comprehensive solution".

"OmicScope .. enabling multi-omics analysi" [line 197] The authors only include proteomics with some transcriptomics, no metabolomics, lipidomics or genomics are addressed. Do the authors intend to define multi-omics as "multiple independent studies"? Some clarity is needed.

Authors claim in the abstract "with over 224 databases"[line 21] without credit to Enrichr until later in the main text. This should be clarified or removed as a unreferenced claim.

The authors state "The complexity and number of computational tools present an initial barrier, particularly for non-programmers and newcomers to proteomics"[line 50] yet fail to address how their implementation specifically benefits this group. The URL for the interactive application most accessible to "non-programmers"[line 51] is included only in the abstract and not mentioned in the main text. Online documentation is sparse, and no cross-references are found between the code, documentation and web app. In addition, the authors fail to address various other softwares for post analysis exploration and visualization such as Pyteomics, Ursgal, pyOpenMS, AlphaMap, MSPypeline, IdentiPy, Scavager, Proteomatic, DEqMS, DAPAR, ProteoSign, tidyproteomics, tidymass, pmartR, protti or recognize their various web-based implementations. The audience would benefit from a discussion and comparison of OmicScope to what is currently available.

The statement " autonomously conducts statistical analysis" [line 275] is an interesting choice for analysis, unwitting and novice users might assume consistent, even identical, statistical applications only to find that is not the case. OmicScope should include some accounting method post application that informs the user how exactly the data was treated so such details can be conveyed and independently replicated.

Authors make the claim "robust analysis" and "robust platform" in 5 instances without defining

what is meant by “robust” and how that attribute of their software makes it a better choice.

The description “.. we expanded our analysis .. introduced two distinct concentrations of Yeast digest ..” [line 330]. It is initially unclear if the SARS-CoV-2 was expended to include the Yeast data, which it should not, please clarify.

OmicScope claims “comprehensive solution for quantitative proteomics data” however, the reviewer could not find methods for data normalization, imputation or methods for outlier rejection or assessment of quantitative accuracy or power. Authors should address these deficiencies or narrow the scope of claims.

The author aptly point out that ”Besides providing a cleaner representation of the network, this strategy also simplifies information extraction, reduces data redundancy without omitting any data, and aids in selecting targets for further experimental validation (Figure 3F).” [line 401] Perhaps this should be emphasized earlier on, authors might consider “simplified network analysis” over comprehensive.

“Two other systems biology strategies employed by Nebula to assess the similarity between studies in a pairwise fashion are similarity analysis and Fisher's exact test.” [line 497] Can the user select other tests such as T.Test, Wilcoxon, or Kolmogorov-Smirnov?

A major omission; “However, where OmicScope particularly shines is in longitudinal analyse” [line 580] No clear demonstration of this was provided. One needs to be provided and discussed or revised to omit the claim.

Reviewer #1 (Remarks on code availability):

Used the online web version which was modern and responsive. I had issues importing some basic datasets with only cryptic error messages. While it did provide lots of colorful plots, many of them were without description of use or significance. A novice researcher might just look for the prettiest plot and skip over something of significance.

Could not demo the Crunfli dataset from the publication, the website limits file size. I could not demo the PXD046709 data set, despite constricting a pdata.xlsx data file, the import was unable to proceed. A third MaxQuant dataset file to import with KeyError: "['Gene names'] not in index". I was only able to demo with the provided example MaxQuant data, however the process is no well documented online. I was looking for some summary information on the dataset contents, samples, replicates etc. I could not perform a PCA without selecting a control other than CoV. Despite there being 3 sample groups, the Barident plot did not show the difference from the two groups against the control. All the plots need descriptions as to what exactly is being plotted, aside from “box plot”. Some parameters lack descriptions, such as “Line” in the hierarchical clustering, and while “Method” did have a tool-tip, it was uninformative, parameters should have a small discussion around each choice. The k-means clustering, while interesting, gave no context for what was being displayed - a researcher needs to explain their findings. The table from the Enrichment Analysis did not provide explanations and significance for each column. The Number of Dysregulation plot did not explain its significance or what “size” is.

Reviewer #2 (Remarks to the Author):

The research article “OmicScope unravels systems-level insights from quantitative proteomics data” from Reis-de-Oliveira and colleagues presents the usage of OmicScope as a tool for quantitative proteomics.
The authors explain the need for new software applications to improve research effectivity in proteomic analysis due to challenges like multiple software tools that have to be applied to cover parts like raw data processing, protein identification/quantitation, enrichment analysis and pathway analysis and the bridging of other omics like transcriptomics or metabolomics.

The topic is important and of interest. The authors provide a good overview of the software and its practical application options.
The illustrations are suitable and the language quality is appropriate.

However, I have several relevant points to note:

-The need to use the software is described and accepted as a general consensus. It would be desirable to provide evidence of this, perhaps through scientific references and examples and indications of the extent to which the software can specifically help here. Is it about speeding up processes? Cost savings?

-The authors should name more precisely the aim of the manuscript. Is this a study that wants to show the feasibility to use the software as an effective analyzing tool?

-Overall, the work gives the impression of, at least in part, advertising the product. No limitations are mentioned and no competing products are mentioned.
In particular, the discussion should be revised with more neutral language.

-There is a lack of information showing how the software is financed and to what extent the authors and their institutions are financially involved. Are there contacts with industry? Has the software already been used in another facility? It would be necessary to provide more precise information on this.

-Limitations are not addressed. Which quantitative proteomics challenges cannot be solved by the software? Is there a possible risk of bias? Does the software always work objectively or can the algorithms only provide a very one-sided picture? How vulnerable is the software to manipulation?

-The authors use proteome data on CoVID-19 as an example to run the provided system.
The available amount of data and the knowledge on CoVID-19 is quite high due the extraordinary position of this virus during the pandemic.
Maybe working with another dataset, from a not so well known disease could be more representative for the proposed overall benefits of the software.

-In my opinion, the “results” section is not really a part that presents results of an analysis.
The manuscript lacks a direct comparison of the software with another tool, analyzing the same data set.

-The manuscript would benefit from examples, addressing practical problems and limitations of typical proteomics approaches. As an example, the mentioned shotgun proteomics method has some limitations like a limited dynamic range with highly abundant proteins possibly overshadowing low-abundance proteins and challenges in the reproducibility of the results.
Is there any advantage that the software can offer here?
A possible example could be the database support that the software can provide. Inaccurate and incomplete databases can induce false-positive or false-negative identifications. Maybe the authors can add such aspects to the discussion section.

In summary if would strongly recommend to revise the manuscrtipt, including a restructuring and changing the focus to a more scientific view of the product in the context of other software applications.

Reviewer #2 (Remarks on code availability):

The files provided are sufficient to independently reproduce the data in a practical approach.

Reviewer #3 (Remarks to the Author):

The work presents OmicScope focusing on quantitative proteomics data analysis with static and longitudinal designs, including OmicScope, EnrichmentScope, and Nebula components. Its applicative Python package and web forms function users with efficiency to access abundant database and to customize outputs. Large-scale protein-protein interactions allows comprehensive analysis of proteomics, genomics, and transcriptomics. Overall, this work is interesting and can be thought relevant to the field. I recommand publication in Nat. Commun. Some comments are listed for author.

1. The advantage and progress of OmicScope still should be specifically described when compared with other software tools, which are probably highlighted in Figures or listed in charts.
2. There are also some unclear statements, like 575 line in Discussion part as 'differential proteomicscomparisons'.
3. Supplementary Figure 4 should be provided with higher reslution.
4. Title format in references 20, 21, 25, 27, 28, 29, 30, 32, 39, 40 should be consistent with others.

REVIEWER COMMENTS

General note: We have numbered all reviewer comments consecutively to make this document easier to navigate

REVIEWER #1:

We thank the reviewer for taking the time to assess our submission and for the constructive feedback. We focus our efforts in response to each specific comment, which we believe covers all issues highlighted in the summary section.

Synopsis:

OmicScope does facilitate proteomics analyses, making some complex methods simpler. There are no new implementations, the full field of similar tools has not been adequately addressed, and the demonstrations are lacking compelling comparisons to what was previously published (e.g was something confirmed or newly discovered). With added clarity and revisions this reviewer might accept publishing.

Summary:

OmicScope appears to be a well formulated python implementation for post analysis of proteomic data with both a programming code base and web implementation. Of particular interest is the multi-omics and longitudinal analyses, along with the demonstrated analysis of available open access data. However, the authors fail to demonstrate a longitudinal analysis, and the multi-dataset demonstration did not highlight new insights, something that would be compelling for the readers.

In addition, the authors fail to address in the introduction, the myriad other software implementations in python and other languages, or define distinguishing characteristics of their implementation from what is already available among a congested field of proteomic analysis tools. Instead leaving a partial survey of similar tools for the discussion. OmicScope applies well worn analyses to quantified proteomics data, however, it offers no quality assessment, manual filtering capabilities, or data treatments, therefore their claim as a “comprehensive solution” seems unwarranted. The authors are advised to narrow their claims, or demonstrate more complete functionality.

Specifics:

1. Authors claim "OmicScope emerges as a comprehensive solution for quantitative proteomics data analysis”[line 18], claiming inputs from a handful of platforms, however, they fail to address widely utilized proteomics platforms such as ProteomeDiscoverer, FragPipe and OpenMS and instead offer a only a generalized input function. Authors should address these platforms, and offer some path to availability to claim “comprehensive solution”.

We appreciate the insightful comments from the reviewer regarding the input format comprehensiveness of OmicScope.

In response to the reviewer's suggestion, we have expanded the range of data formats supported by our open-source platform. Specifically, we have incorporated support for Proteome Discoverer and FragPipe formats. This expansion demonstrates our commitment to evolving with the community's needs and ensuring that OmicScope remains a versatile tool for proteomics data analysis.

Additionally, we are actively working to integrate support for the OpenMS framework outputs in future updates of OmicScope. We noticed OpenMS exports quantitative information

based on peptide level, requiring additional steps to accurately infer protein quantitation, which still is not covered by OmicScope.

As highlighted in the revised supplementary Table 1, OmicScope now supports a versatile array of data formats in the proteomics field, underscoring our platform's adaptability and utility. These improvements have been inserted in the manuscript (lines: 152-155, 282-284, 609-615), in Figure 1, Figure 2A, Supplemental Table 1, and the Appendix (lines 39-49).

2. “OmicScope .. enabling multi-omics analysi” [line 197] The authors only include proteomics with some transcriptomics, no metabolomics, lipidomics or genomics are addressed. Do the authors intend to define multi-omics as “multiple independent studies”? Some clarity is needed.

Thank you for pointing out the need to clarify our use of "multi-omics" in the manuscript. We agree that OmicScope focuses primarily on proteomic datasets, and our example was less extended to a few elements of transcriptomics. However, despite Nebula being designed for meta-analysis purposes, our results demonstrate OmicScope’s versatility while handling multiple types of omics data.

Additionally, other features in OmicScope extend its functionality beyond proteomics. The "General Import" method allows for differential expression, proteomics, and abundance analyses. Although our examples focus on proteomics, this method can accommodate other mass-spectrometry-based omics data, such as metabolomics and lipidomics, by enabling group comparisons based on abundance levels. Additionally, the "Snapshot" import method enables users to import previously analyzed data, extending its scope to include genomics and transcriptomics analyses and enhancing OmicScope's applicability.

Regarding enrichment analysis, OmicScope uses "gene name" as a parser to match with relevant target databases. This approach allows EnrichmentScope to perform enrichment analyses with proteomics, genomics, and transcriptomics data.

Finally, the Nebula module integrates information from OmicScope and EnrichmentScope, allowing users to concatenate and compare datasets previously imported into OmicScope workflow. In our manuscript, the Nebula module was used to compare proteomic and transcriptomic datasets, showcasing OmicScope's capabilities to perform meta-analyzes beyond proteomics experiments.

To clarify the above topics in our manuscript, we avoided “multi-omics” expression and have added information in lines 201-202 and 214-217, explicitly demonstrating that Nebula is a module dedicated to performing meta-analysis, despite its functions being not limited to proteomics experiments.

3. Authors claim in the abstract “with over 224 databases”[line 21] without credit to Enrichr until later in the main text. This should be clarified or removed as an unreferenced claim.

Thank you for pointing out the omission of credit to Enrichr in the abstract. We recognize the importance of properly referencing sources and have updated the abstract to include proper credit to Enrichr in line 22. Additionally, we have added a citation in Figure 1 to ensure Enrichr is credited in the visual representation of our work.

4. The authors state “The complexity and number of computational tools present an initial barrier, particularly for non-programmers and newcomers to proteomics”[line 50] yet fail to address how their implementation specifically benefits this group. The URL for the interactive application most accessible to “non-programmers”[line 51] is included only in the abstract and not mentioned in the main text.

We appreciate the reviewer's comment on addressing the benefits of our implementation for non-programmers and newcomers to proteomics. The initial manuscript contained limited information about how our approach specifically supports this group, and we have taken steps to resolve this.

To address this concern, we have provided more detailed information about the web app in the Results (lines 261-250) and Discussion (lines: 644-647, 688-694) section, mainly highlighting user-friendly and interactivity features associated with OmicScope app.

Additionally, to make our application more accessible, we have included the URL in all sections, ensuring readers can easily locate and use the tool.

5. Online documentation is sparse, and no cross-references are found between the code, documentation and web app.

Thank you for the reviewer's observation about the sparse online documentation and the lack of cross-references between the code, documentation, and web app.

To address this feedback, we improved the quality and coherence of our documentation, improving experience for users across all platforms. Initially, we published the online documentation, including step-by-step instructions to analyze data using OmicScope Python package, with usage examples, on https://OmicScope.readthedocs.io/en/. We have added in the manuscript the URL for online documentation (lines 119-120).

Additionally, to improve cross-references, we have established links between the web app and the online documentation. This interlinking ensures that users can easily navigate between different sources of information, allowing them to understand the context and workflow better. Finally, we also provide video tutorials in the OmicScope application to guide users in their first steps, mentioning both sources of information.

6.In addition, the authors fail to address various other softwares for post analysis exploration and visualization such as Pyteomics, Ursgal, pyOpenMS, AlphaMap, MSPypeline, IdentiPy, Scavager, Proteomatic, DEqMS, DAPAR, ProteoSign, tidyproteomics, tidymass, pmartR, protti or recognize their various web-based implementations. The audience would benefit from a discussion and comparison of OmicScope to what is currently available.

We appreciate the reviewer's inquiry about comparing OmicScope to other software tools in the field of proteomics analysis. To provide a comprehensive comparison, we conducted an extensive investigation of 23 computational tools, including those suggested by the reviewer (Supplementary table 1).

Using a filter to select tools that enable at least differential proteomics analysis, we identified several tools that were not directly comparable to OmicScope due to their focus on different aspects of proteomics analysis (such as identification and quantitation) or their lack of protein-level analysis capabilities. These excluded tools include AlphaMap, Identipy, Scavager, Proteomatic, tidymass, Ursgal, pyOpenMS, and Pyteomics.

For the remaining tools that were comparable to OmicScope in scope, we compiled Supplementary Table 1, which compares each software based on various criteria including import methods, parameters for differential proteomics analysis, enrichment capabilities, meta-analysis functionalities, and export methods. Through this comparison, we aimed to highlight the features and capabilities of OmicScope against other available tools. This information and survey was added to the main text (lines 229-237).

In our survey, OmicScope emerged as a versatile platform that covers several aspects of the proteomics workflow, including functionalities not implemented or encompassed by other platforms. our results indicate that most tools lack the capability for multiple group comparisons or integration of data from different sources, such as enrichment analysis and differential proteomics data. Besides EnrichmentScope integrating quantitative and enrichment outcomes, the Nebula module allows multi-study comparisons at protein- and enrichment-level.

Additionally, while looking for software that presents a broad analysis, such as Perseus, we noted that even those tools present limitations. Perseus, for instance, does not include databases for performing protein-protein interaction networks or enrichment analyses, requiring users to manually download and integrate these resources into their workflow. On the other hand, OmicScope is empowered by Enrichr and provides more than 224 libraries to perform enrichment analysis (ORA and GSEA).

Finally, certain tools are only available as programming language packages, which may require a minimum level of programmatic skills to carry out analysis. OmicScope is available as a Python package and web application, allowing easy utilization by both experienced programmers and newcomers in proteomics.

We added these comparisons and respective discussions throughout the main text (Result and Discussion sections), referencing supplementary table 1.

7. The statement " autonomously conducts statistical analysis" [line 275] is an interesting choice for analysis, unwitting and novice users might assume consistent, even identical, statistical applications only to find that is not the case. OmicScope should include some accounting method post application that informs the user how exactly the data was treated so such details can be conveyed and independently replicated.

We appreciate the reviewer's insight regarding the need for a post-application accounting method to ensure users can understand and replicate the analysis performed by OmicScope. To address this, we have implemented a new object, OmicScopeObject.Params, in the OmicScope package. This object keeps a record of all actions taken during an analysis, including import, pre-processing, filtering, and statistical steps. Using this object, users can generate a report that summarizes the statistical methods applied during their analysis.

Additionally, within the OmicScope App, we have included a table that contains information by this object, allowing users to download it for further reference or documentation.

To clarify the functionality of OmicScopeObject.Params, we have added explanations in the Appendix, specifically on lines 94-96 and 136-137.

8. Authors make the claim "robust analysis" and "robust platform" in 5 instances without defining what is meant by "robust" and how that attribute of their software makes it a better choice.

Thank you for pointing out the ambiguity in our use of the term "robust". After reviewing the manuscript, we agree that this term was used in multiple instances without a proper definition or explanation of its relevance to our software's performance.

On line 358, we used "robust" to describe the performance of OmicScope against a benchmark dataset containing over 12,000 proteins, which included differentially regulated proteins from the yeast proteome. To improve clarity, we replaced "robust" with "reproducible," emphasizing the consistency and reliability of the results, since the results were similar to those previously published by Demichev and Meier. This change aligns better with the context.

In the legend for Figure 2, on line 591, and on line 648, we omitted the term "robust" to prevent confusion. On line 630, where we mentioned "robust results," we replaced it with “consistent outcomes”, which aligns better with the context.

9. The description “.. we expanded our analysis .. introduced two distinct concentrations of Yeast digest ..” [line 330]. It is initially unclear if the SARS-CoV-2 was expended to include the Yeast data, which it should not, please clarify.

We appreciate the reviewer's comment and recognize that our previous description may have caused confusion. To clarify, the section in question refers to a demonstration of OmicScope’s ability to handle large datasets and identify differentially regulated proteins. This experiment involves introducing two distinct concentrations of yeast proteome into HeLa digest, with SARS-CoV-2 not being part of this dataset.

In Supplementary Figures 2A and 2B, we illustrate how OmicScope successfully analyzed the Yeast dataset with over 12,000 proteins, determining a subset of known differentially regulated proteins. Data highlighted in red (Supplementary Figure 2B) comprises the fold-change between both concentrations of yeast proteome spiked into the HeLa digest. These results align with the findings of the original publication from Meier and Demichev.

We acknowledge that the initial wording could suggest a connection to SARS-CoV-2, which is incorrect. We clarify this issue in the manuscript (lines 364-375) to ensure that the description accurately reflects the purpose and outcome of the experiment.

10. OmicScope claims “comprehensive solution for quantitative proteomics data” however, the reviewer could not find methods for data normalization, imputation or methods for outlier rejection or assessment of quantitative accuracy or power. Authors should address these deficiencies or narrow the scope of claims.

Thank you for pointing out the potential overstatement in our claim of "comprehensive solution for quantitative proteomics data".

After revising the manuscript, we agree that the term "comprehensive" may suggest a broader scope than we intended, therefore we have rewritten the abstract on lines 19-21 to reflect a more precise scope. Additionally, to avoid overstatement, we have omitted the word "comprehensive" from lines 116, 392, and 640. These changes align our claims with the current features and focus of our software, providing a more accurate representation of its capabilities.

On the other hand, based on the reviewer’s suggestions and to demonstrate our commitment to meeting the demands of the scientific community, we also have implemented two new features in the latest version of our software: Data Normalization and Imputation. Here, we introduced three different methods for data normalization and imputation, allowing users to preprocess their data before statistical analysis. These features are available in both the Python package and the web application, providing flexibility for users with different workflows. These implementations were incorporated in the manuscript (on lines 21, 164-168, 305), Figure 1A, and Appendix (on lines 91-92, and 111-114).

11. The author aptly point out that ”Besides providing a cleaner representation of the network, this strategy also simplifies information extraction, reduces data redundancy without omitting any data, and aids in selecting targets for further experimental validation (Figure 3F).” [line 401] Perhaps this should be emphasized earlier on, authors might consider “simplified network analysis” over comprehensive.

We appreciate the reviewer's observation regarding the emphasis of the point about simplified network analysis. To address this, we have revised the text to highlight the benefits of this approach earlier in the paragraph, specifically starting from line 445.

12. “Two other systems biology strategies employed by Nebula to assess the similarity between studies in a pairwise fashion are similarity analysis and Fisher's exact test.” [line 497] Can the user select other tests such as T.Test, Wilcoxon, or Kolmogorov-Smirnov?

Thank you for the reviewer's question regarding the statistical tests used by Nebula to compare studies in a pairwise fashion. Initially, Nebula only performed Fisher's exact test, which is commonly used for enrichment analysis (over-representation analysis) to determine whether a subset of proteins is over-represented in another set, considering a background.

However, in response to the reviewer’s feedback and our commitment to evolving OmicScope based on user needs, we have implemented additional statistical tests in the current version of Nebula. Users can now select from the following tests for comparing studies:

1. T-test: This test allows you to compare the means of two groups to determine if there is a statistically significant difference between them.
2. Wilcoxon Test: A non-parametric test used to compare two paired groups, suitable for data that does not follow a normal distribution.
3. Kolmogorov-Smirnov Test: This test compares the distribution of two independent datasets to determine if they differ significantly.

With these new options, Nebula can analyze proteins that overlap between pairwise studies and compare the fold-change distribution of proteins (typically focusing on differentially regulated ones). The default setting assumes the null hypothesis that the fold-change distributions are the same between groups. While using these tests, the new stat_network/stat_heatmap function (which replaces fisher_network/fisher_heatmap) establishes a link between groups if the p-value is greater than or equal to 0.05, indicating that the distributions are not significantly different.

We adapted our manuscript on line 553-568, figure 5, and appendix (257-266).

13. A major omission; “However, where OmicScope particularly shines is in longitudinal analyse” [line 580] No clear demonstration of this was provided. One needs to be provided and discussed or revised to omit the claim.

Thank you for pointing out the omission in our manuscript regarding the demonstration of OmicScope's capabilities in longitudinal analysis. We appreciate the opportunity to provide clarification and evidence to support our claim.

To address this, we utilized OmicScope to evaluate a dataset published by Grossegesse (analysis in figure below), which investigated proteome changes triggered by SARS-CoV-2 infection compared to a Mock infection (control) at four-time points (2, 6, 10, and 12 hours post-infection), on CaLu cell lines.

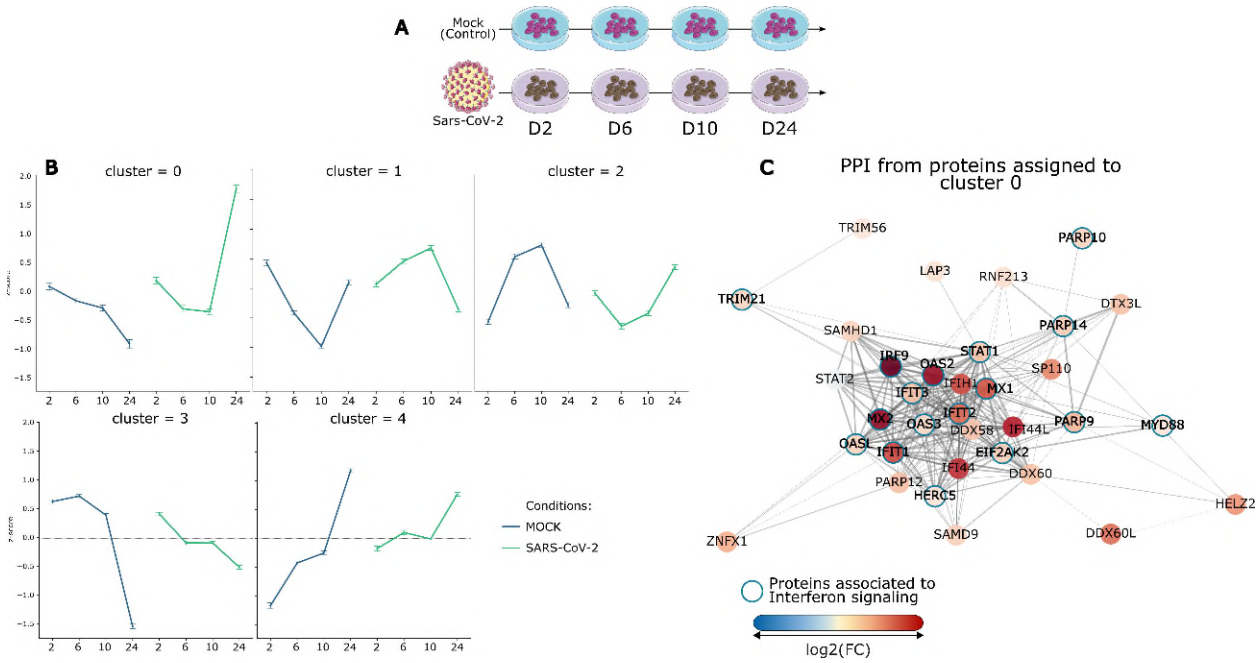
Through OmicScope analysis of the dataset, we identified 614 differentially regulated proteins (DRPs), using between-class comparison. To evaluate changes in DRP patterns over

time, we applied a K-trend plot analysis, which revealed five clusters. Among these clusters, three exhibited changes in abundance patterns in opposite directions.

Further investigation focused on the cluster with the highest fold-change variation (cluster 0). A protein-protein interaction search was conducted to identify potential links among these proteins. We found that several proteins in this cluster were associated with the interferon pathway and showed up-regulation throughout SARS-CoV-2 infection, as previously discussed in literature.

These results demonstrate OmicScope's capacity to handle longitudinal experimental designs and its integrative capabilities in performing clustering analysis and protein-protein interaction searches within a single environment.

We have included this analysis in the main text (on lines: 376-390) of the manuscript and have added supplementary figure 3 to provide visual support for the findings. Additionally, we also have omitted the sentence "However, where OmicScope particularly shines is in longitudinal analyse" to avoid overstatement and improve the scientific tone.



Supplementary Figure 3. ***Longitudinal Analysis.*** *A)* Grossegesse conducted longitudinal analysis to compare the effect of SARS-CoV-2 infection on CaLu cell lines, contrasting this response against a control group (Mock). B) We identified 614 differentially regulated proteins, which were subjected to K-means clustering to identify distinct abundance patterns over time. In this analysis, we identified 5 clusters, with clusters 0, 1, and 2 exhibiting the most distinct patterns. C) Using proteins assigned to cluster 0, we conducted a protein-protein interaction search using the OmicScope algorithm. Our survey revealed an upregulated protein cluster associated with interferon signaling.

Remarks on code availability:

14. Used the online web version which was modern and responsive. I had issues importing some basic datasets with only cryptic error messages. While it did provide lots of colorful plots, many of them were without description of use or significance. A novice researcher might just look for the prettiest plot and skip over something of significance.

We appreciate the reviewer's feedback on the user experience with our online web version. Although the web app is designed to be modern and responsive, we understand that cryptic error messages and the lack of plot descriptions could lead to confusion, particularly for novice researchers.

To address these issues, we have revised the error messages to provide more detailed information on common import issues. This should help users understand the cause of an error and guide them in correcting it, reducing frustration and improving the data import process.

In addition, we have updated the online documentation on the ReadTheDocs website, providing explanations for each chart type generated by the OmicScope Package. In the web app, we’ve added a "How to Interpret" drop-down menu next to each chart, explaining its use and meaning. This feature aims to help users better understand their data by ensuring that they do not ignore important information due to unclear descriptions.

With these changes, we hope to improve the user experience for both novice and experienced researchers, reducing the likelihood of misinterpretation and enhancing overall usability.

15. Could not demo the Crunfli dataset from the publication, the website limits file size. I could not demo the PXD046709 data set, despite constricting a pdata.xlsx data file, the import was unable to proceed. A third MaxQuant dataset file to import with KeyError: "['Gene names'] not in index". I was only able to demo with the provided example MaxQuant data, however the process is no well documented online. I was looking for some summary information on the dataset contents, samples, replicates etc.

We apologize for the inconvenience experienced while importing data into OmicScope on the web application. We understand that these issues can be frustrating, and we have taken steps to address each concern.

1. Crunfli Dataset: We have ensured that the Crunfli dataset is available on our website and as supplementary material. This dataset is smaller than 250 MB, so it should be importable without issues. However, to resolve any confusion, in addition to supplementary table 2, we've provided a Google Drive link with the correct data for easy access:

   https://docs.google.com/spreadsheets/d/1I_HP7QQl2fYAtMEmoS5X99Gf2EbAKePg/edit?usp=drive_link&ouid=107564255465897822066&rtpof=true&sd=true

   Additionally, to prevent file size-related issues, we have increased the maximum file size for statistical analysis to 3 GB.

2. PXD046709 Dataset: OmicScope application uses 'LFQ Intensity' as the quantification strategy recommended by MaxQuant's team. However, the PXD046709 dataset uses raw "Intensity" values instead of 'LFQ Intensity,' leading to the reported error. We have corrected this problem by allowing OmicScope to use 'Intensity' when 'LFQ Intensity' is not available and we improved the error messages. We apologize for the confusion and thank you for pointing out the issue.

Additionally, upon further investigation, we found that the PXD046709 dataset lacks quantitative values, preventing OmicScope from performing statistical analysis (image below).

CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO
Intensity 007_1	Intensity 007_2	Intensity 007_3	Intensity 007_4	Intensity 6MB_1	Intensity 6MB_2	Intensity 6MB_3	Intensity 6MB_4	Intensity ctrl1	Intensity ctrl2	Intensity ctrl3	Intensity ctrl4
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

3. KeyError: "['Gene names'] not in index": We investigated this error but could not reproduce it. It is possible that the imported data lacked the "Gene names" column or used a different version of MaxQuant. To address this, we have implemented additional checks and improved the documentation to help users troubleshoot similar issues.

Additionally, we have updated our online documentation to provide clearer instructions on the data import process and added more content to our website. This includes additional guidance on dataset contents, samples, replicates, and other relevant information to assist users in navigating the web application.

These updates should resolve the problems mentioned and improve the user experience with OmicScope.

16. I could not perform a PCA without selecting a control other than CoV.

We apologize for the inconvenience you experienced while attempting to perform PCA in the OmicScope app. This issue occurred because the app required users to select a control group, with "CoV" being the default. We recognize that this constraint limited the flexibility and usability of the PCA function.

To resolve this issue, we have implemented a fix in the OmicScope app. Now, when users import their data, the app displays all groups present in the dataset, allowing users to select the appropriate control group for their PCA analysis. This update ensures that users can perform PCA with greater flexibility, selecting the desired groups and controls without restrictions.

17. Despite there being 3 sample groups, the Baridentplot did not show the difference from the two groups against the control. All the plots need descriptions as to what exactly is being plotted, aside from “box plot”. Some parameters lack descriptions, such as “Line” in the hierarchical clustering, and while “Method” did have a tool-tip, it was uninformative, parameters should have a small discussion around each choice. The k-means clustering, while interesting, gave no context for what was being displayed - a researcher needs to explain their findings. The table from the Enrichment Analysis did not provide explanations and significance for each column. The Number of Dysregulation plot did not explain its significance or what “size” is.

We appreciate the reviewer's valuable comments, which highlight the need for clearer plot descriptions and better explanations of various parameters.

Regarding the barident plot, which aims to show a high-level overview of differentially regulated proteins, can be confusing if it does not specify differences among groups. Although the current version focuses on a general overview, we plan to improve the barident plot to provide more specific information about the number of differentially regulated proteins across different groups.

Additionally, we have added more descriptive text for each plot to help users understand what is being displayed. This includes specific explanations for common plots like "box plots," "k-means clustering," and others. The "How to Interpret" drop-down menu within the web application has been used to help users with a more detailed guidance on each plot (and Enrichment Table), offering researchers a better understanding of their data.

Finally, parameters such as "Line" in hierarchical clustering and "Method" have been given more detailed tooltips, explaining what they represent and how they affect the analysis. This should help users make informed choices when configuring their analyses.

REVIEWER #2:

We thank the reviewer for their constructive criticism and hope we have adequately addressed their comments.

The research article “OmicScope unravels systems-level insights from quantitative proteomics data” from Reis-de-Oliveira and colleagues presents the usage of OmicScope as a tool for quantitative proteomics.
The authors explain the need for new software applications to improve research effectivity in proteomic analysis due to challenges like multiple software tools that have to be applied to cover parts like raw data processing, protein identification/quantitation, enrichment analysis and pathway analysis and the bridging of other omics like transcriptomics or metabolomics.

The topic is important and of interest. The authors provide a good overview of the software and its practical application options.
The illustrations are suitable and the language quality is appropriate.

However, I have several relevant points to note:

18. The need to use the software is described and accepted as a general consensus. It would be desirable to provide evidence of this, perhaps through scientific references and examples and indications of the extent to which the software can specifically help here. Is it about speeding up processes? Cost savings?

Thank you for highlighting the lack while providing evidence for the utility of OmicScope in proteomics data analysis. In the field of mass spectrometry-based proteomics, researchers face numerous challenges regarding data analysis, which include the numerous software required to perform: protein identification and quantitation, differential proteomics analysis, search for protein-protein interactions, enrichment analysis, meta-analysis. In addition, the integration among those platform also is a challenging, since the field presents variations in data formats exported by each software and lacks integrative tools to perform downstream analysis. OmicScope addresses these challenges by offering solutions for data handling, normalization, differential proteomics analysis, enrichment analysis, and meta-analysis, all within a single computational environment.

By simplifying these processes and providing a user-friendly interface, OmicScope accelerates data analysis, reduces the time required for manual tasks such as table preparation, and allows users to gain insights at multiple levels of analysis, from individual proteins to enriched pathways and meta-analyses across multiple studies. This not only enhances research productivity but also facilitates accessibility for non-coding users and newcomers to proteomics.

To support these claims, we referenced scientific literature (references 9-11) that discusses the challenges faced in proteomics research. Additionally, we add throughout the manuscript how OmicScope faces the several challenges (examples on lines: 251-265, 299-301, 304-307, and Discussion section).

19. The authors should name more precisely the aim of the manuscript. Is this a study that wants to show the feasibility to use the software as an effective analyzing tool?

Thank you for your feedback on clarifying the aim of the manuscript.

The primary aim of our study is to introduce OmicScope as a novel and versatile tool for proteomic data analysis. OmicScope offers an integrated environment that encompasses key aspects of the proteomics workflow, including differential proteomics, enrichment analysis, meta-analysis, which integrates various omics datasets.

Through our manuscript, we aim to demonstrate how OmicScope facilitates the proteomic data analysis process, providing users with a standardized approach to extract insights from protein-level data to systems-level understanding. Additionally, we highlight OmicScope's ability to integrate with upstream and downstream tools, offering a versatile solution for researchers in the field of proteomics. Throughout manuscript we analyze data and suggest potential biological insights that OmicScope could provide to users.

We have revised the manuscript to provide clearer language regarding the specific aims of our study, as indicated in lines 95-111, 238-240, 595-600.

20. Overall, the work gives the impression of, at least in part, advertising the product. No limitations are mentioned and no competing products are mentioned.

We appreciate the reviewer's suggestion to include a comparison with other computational tools related to OmicScope. We have addressed this concern by adding Supplementary Table 1, which compares 15 software tools for downstream analysis of proteomics data, including OmicScope. This table provides a comprehensive comparison based on various criteria such as input format acceptance, statistical workflow, enrichment analysis capabilities, meta-analysis features, and export formats. Additionally, in the discussion section, we have provided a more direct comparison among computational tools, highlighting their respective advantages and limitations.

Furthermore, we acknowledge the reviewer's observation regarding the lack of mention of limitations in the manuscript. We have addressed this concern by discussing OmicScope's limitations in response to a question raised by the same reviewer (question number 23). We have provided the potential limitations of OmicScope, including the risk of bias, overfitting, and the software's vulnerability to manipulation, while also outlining its ongoing evolution to address additional quantitative challenges in proteomics research.

21. In particular, the discussion should be revised with more neutral language.

Thank you for bringing this to our attention. We revised the discussion section to ensure that the language used is more neutral and precise.

22. There is a lack of information showing how the software is financed and to what extent the authors and their institutions are financially involved. Are there contacts with industry? Has the software already been used in another facility? It would be necessary to provide more precise information on this.

Thank you for your question regarding the financing and institutional involvement related to the development of the software. The software was developed by the authors of the article, with support from FAPESP and CNPq, Brazilian financing agencies. We included the grant numbers provided by these agencies in the text for transparency. Noteworthy, there has been no industrial support or investment in the development of the software and the State University of Campinas (Unicamp) will continue to support the maintenance of the web page and provide the domain for the web tool.

23. Limitations are not addressed. Which quantitative proteomics challenges cannot be solved by the software? Is there a possible risk of bias? Does the software always work objectively or can the algorithms only provide a very one-sided picture? How vulnerable is the software to manipulation?

Thank you for addressing the reviewer's concerns regarding OmicScope's limitations. OmicScope was indeed designed to address the most common quantitative challenges encountered in proteomics data analysis, catering to both novice and experienced users. While the software offers flexibility in parameter adjustments to accommodate various experimental designs, there is a potential risk of bias and overfitting when users modify these parameters extensively. To mitigate this risk, OmicScope provides stringent default parameters for statistical analysis, minimizing data manipulation. By defaulting to these parameters, the software aims to present users with an objective and unbiased picture of the data, while still allowing experienced users the option to make adjustments responsibly.

Furthermore, OmicScope facilitates transparent reporting of the statistical pipeline by allowing users to retrieve steps performed using command line or downloading from web toll. This feature enhances reproducibility and transparency in data analysis, aligning with best practices in scientific research, and can be requested by the reviewers to confirm all the parameters utilized by the user.

Critically, our tool currently lacks methods for outlier rejection, inclusion of additional filtering steps, and assessment of quantitative accuracy, which will be the focus of future implementations. In addition to these implementations, as part of its ongoing evolution, the software intends to incorporate features to address more specific quantitative challenges encountered in proteomics research. These may include handling data from experiments such as dose-response curves, thermal shifting assays, and co-expression analysis, thereby expanding OmicScope's utility and relevance to a broader range of proteomics investigations.

We incorporated a brief discussion on these topics in the main manuscript (lines:627-630, 647-654, 699-701).

24. The authors use proteome data on CoVID-19 as an example to run the provided system. The available amount of data and the knowledge on CoVID-19 is quite high due the extraordinary position of this virus during the pandemic.
Maybe working with another dataset, from a not so well known disease could be more representative for the proposed overall benefits of the software.

Thank you for your valuable feedback regarding the choice of datasets used in our study.

We acknowledge the reviewer's suggestion to consider working with datasets from lesser-known diseases to highlight the broader benefits of OmicScope. While we agree that exploring datasets from other diseases could provide additional insights, we chose to utilize COVID-19 datasets for several reasons.

Firstly, the COVID-19 pandemic has resulted in an unprecedented amount of available data due to the global scientific community's efforts to make data openly accessible. This abundance of data facilitated our evaluation of OmicScope's performance with diverse COVID-19 datasets, spanning different tissues, omics types, and experimental designs.

Furthermore, the unique nature of the COVID-19 pandemic allowed us to access datasets that encompass a wide range of experimental conditions and biological responses

triggered by the virus. By analyzing these datasets, we were able to demonstrate OmicScope's versatility in handling complex and diverse data types and sizes.

In addition, we included an analysis of a benchmark dataset containing differentially regulated proteins from Yeast, demonstrating OmicScope's capabilities beyond COVID-19 datasets and even human-related diseases.

25. In my opinion, the "results" section is not really a part that presents results of an analysis. The manuscript lacks a direct comparison of the software with another tool, analyzing the same data set.

Thank you for your feedback.

We appreciate your suggestion to include a direct comparison of OmicScope with other tools analyzing the same dataset. In response, we have added Supplementary Table 1 to the manuscript, which provides a comprehensive comparison of OmicScope with other existing software tools for proteomics data analysis. This table allows readers to evaluate the features and capabilities of each tool, enabling them to make informed decisions about which tool best suits their specific needs. This table was used throughout main manuscript to compare OmicScope capabilities against other available tools.

Additionally, upon reviewing the table, it becomes apparent that direct comparisons using the same datasets pose challenges due to differences in data inputs, analysis methods, and available functionalities among the tools. While some tools may stand out in certain aspects, they may lack others that are essential for comprehensive proteomics data analysis and comparison.

We believe that by presenting this comparative analysis, readers can gain a clear understanding of OmicScope's unique features and advantages compared to other tools in the field. While direct comparisons using identical datasets may not be feasible, this table serves as a valuable resource for researchers to assess the suitability of OmicScope for their specific research needs.

26. The manuscript would benefit from examples, addressing practical problems and limitations of typical proteomics approaches. As an example, the mentioned shotgun proteomics method has some limitations like a limited dynamic range with highly abundant proteins possibly overshadowing low-abundance proteins and challenges in the reproducibility of the results.
Is there any advantage that the software can offer here?
A possible example could be the database support that the software can provide. Inaccurate and incomplete databases can induce false-positive or false-negative identifications. Maybe the authors can add such aspects to the discussion section.

Thank you for the insightful observations regarding practical problems and limitations associated with shotgun proteomics methods. Indeed, the identification of low-abundant proteins with high confidence and avoiding ion suppression remain significant challenges in the field of proteomics, which we do not dare to resolve with our algorithm.

Nowadays, processing software addresses this by employing artificial intelligence (AI) prediction of spectra, which complements the lack of information with a spectral library, thereby confirming peptide identification through the fragmentation pattern. Our software focuses on post-analysis rather than performing peak search or database search. It utilizes exported files from processing software for data treatment, quality analysis, enrichment analysis, and data integration. In this regard, OmicScope supports files from DIA-NN, Progenesis, and Proteome Discoverer, which employ spectral complementation via artificial

intelligence. Additionally, even for downstream analysis, OmicScope offers several features that can help researchers address these challenges effectively.

Firstly, OmicScope implements filters to identify proteins that are present in at least one sample per condition, helping to mitigate issues related to stochastic features and missing values in proteomics data. This ensures that the analysis focuses on proteins with reliable measurements, even those with lower abundances.

OmicScope also provides tools for normalization and data imputation, enabling users to preprocess their data effectively before performing statistical analysis. This can help to improve the accuracy and reliability of downstream results, particularly for proteins with varying abundance levels.

Additionally, OmicScope offers a range of visualization tools, including correlation plots, dynamic range analysis, and volcano plots. These visualizations allow users to explore the data and identify patterns or outliers that may affect the analysis. Based on these insights, users can perform data preprocessing, outlier removal on original tables, and the selection of appropriate software for identification/quantitation.

In summary if would strongly recommend to revise the manuscrtipt, including a restructuring and changing the focus to a more scientific view of the product in the context of other software applications.

REVIEWER #3:

We appreciate reviewer's positive and constructive assessment of our work.

The work presents OmicScope focusing on quantitative proteomics data analysis with static and longitudinal designs, including OmicScope, EnrichmentScope, and Nebula components. Its applicative Python package and web forms function users with efficiency to access abundant database and to customize outputs. Large-scale protein-protein interactions allows comprehensive analysis of proteomics, genomics, and transcriptomics. Overall, this work is interesting and can be thought relevant to the field. I recommand publication in Nat. Commun. Some comments are listed for author.

27. The advantage and progress of OmicScope still should be specifically described when compared with other software tools, which are probably highlighted in Figures or listed in charts.

Thank you for your feedback. We have incorporated a detailed comparison between OmicScope and other software tools in Supplementary Table 1, which provides an overview of the advantages and progress of OmicScope in relation to other tools. This comparison includes various aspects such as data input formats, analysis capabilities, and visualization features, allowing for a thorough assessment of OmicScope's strengths compared to other software tools. We also referred throughout the manuscript (mainly in Result and Discussion section) this table to compare OmicScope with other computational tools

28. There are also some unclear statements, like 575 line in Discussion part as 'differential proteomicscomparisons'.

Thank you for bringing this to our attention.

We have carefully reviewed the manuscript and have addressed the unclear statement you mentioned in the Discussion section. Sentences like "differential proteomicscomparisons" has been revised for clarity to ensure that it accurately reflects the intended meaning.

29. Supplementary Figure 4 should be provided with higher reslution.

Thank you for bringing this to our attention.

We have updated Supplementary Figure 5 (previous Sup. Figure 4) to provide a higher resolution version for better clarity.

30. Title format in references 20, 21, 25, 27, 28, 29, 30, 32, 39, 40 should be consistent with others.

Thank you for bringing this inconsistency to our attention. We have reviewed the references and ensured that the title format is consistent with others throughout the manuscript.

Reviewer #1 (Remarks to the Author):

Nature Comm. Review

Overall the authors responses to the initial reviewer comments were well addressed with earnest devotion to improving both the manuscript and the software. The most noteworthy results of this manuscript have become clearer, with the software demonstrating an ability to perform several sequential steps involved in the proteomics post processing to gain insight to differential expression, gene pathway/network activation, and multiomic biological activity. As the authors point out, there are several software tools available that fulfill a single or partial analysis, OmicsScope attempts to provide a complete pipeline.

The reviewer recommends publishing with only a minor additional revision suggestion.

A single minor corrections.
1. Supplementary Table 1 should also detail data pretreatment steps of Normalization, Imputation and Filtering. The reviewer believes this will further differentiate OmicsScope.

Reviewer #1 (Remarks on code availability):

The code is available as both a python package and web application. The online documentation has been extensively revised. The web application has been updated to provide tool tips, better error support and more informative plot descriptions.

Reviewer #2 (Remarks to the Author):

The authors revised the manuscript extensively.
My comments focused on the inclusion of scientific references and examples and indications to illustrate potential benefits and usability.
Further, I suggested to include information about similar software tools and funding.
They addressed all of these aspects very well.
Adding the differentially regulated proteins from Yeast is beneficial for the manuscript.
The revised discussion section now includes parts on limitations,
In summary, the authors addressed all my raised points appropriately.
The answers to the comments from reviewer #1 are precise and helpful, especially the code availability and the addition of Supplementary Figure 3.
The manuscript improves a lot, so no further comments from my side.

Reviewer #2 (Remarks on code availability):

The authors revised the error messages and updated the online documentation.
Demo the Crunfli Dataset is now possible.
PCA analysis is more flexible now.

REVIEWER #1:

Remarks to the Author

Overall the authors responses to the initial reviewer comments were well addressed with earnest devotion to improving both the manuscript and the software. The most noteworthy results of this manuscript have become clearer, with the software demonstrating an ability to perform several sequential steps involved in the proteomics post processing to gain insight to differential expression, gene pathway/network activation, and multiomic biological activity. As the authors point out, there are several software tools available that fulfill a single or partial analysis, OmicsScope attempts to provide a complete pipeline.

We thank the reviewer for taking the time to evaluate OmicScope and for their constructive feedback. We are pleased that our efforts to address the initial comments have clarified the manuscript's most noteworthy results and demonstrated the software's capabilities. Your recognition of OmicScope's improvements and completeness for proteomics post-processing is greatly appreciated.

The reviewer recommends publishing with only a minor additional revision suggestion.

A single minor corrections.

1. Supplementary Table 1 should also detail data pretreatment steps of Normalization, Imputation and Filtering. The reviewer believes this will further differentiate OmicsScope.

We thank the reviewer for their attention to this detail. We have added an additional section to Supplementary Table 1, which now includes information on pre-processing steps carried out by the evaluated tools. This section details their abilities and methods for normalization, data imputation, data transformation, and filtering steps.

Remarks on code availability

The code is available as both a python package and web application. The online documentation has been extensively revised. The web application has been updated to provide tool tips, better error support and more informative plot descriptions.

REVIEWER #2

Remarks to the Author

The authors revised the manuscript extensively.
My comments focused on the inclusion of scientific references and examples and indications to illustrate potential benefits and usability.
Further, I suggested to include information about similar software tools and funding.
They addressed all of these aspects very well.
Adding the differentially regulated proteins from Yeast is beneficial for the manuscript.
The revised discussion section now includes parts on limitations,
In summary, the authors addressed all my raised points appropriately.
The answers to the comments from reviewer #1 are precise and helpful, especially the code availability and the addition of Supplementary Figure 3.
The manuscript improves a lot, so no further comments from my side.

Remarks on code availability

The authors revised the error messages and updated the online documentation.

Demo the Crunfli Dataset is now possible.
PCA analysis is more flexible now.

We appreciate the reviewer's feedback and the time taken to help us improve our tool. Your comments were instrumental in enhancing the manuscript and the software.