nature portfolio

Peer Review File

High-Throughput Label-free Single-Cell Proteomics Enabled by Multicolumn NanoLC

Corresponding Author: Dr Ryan Kelly

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this fascinating and scientifically interesting manuscript, the authors describe a high-throughput multicolumn LC system that is a valuable addition to the field of SCP, helping to evolve its throughput to the next level and transition the field from method development to the biological toolbox.
By enabling the routine analysis of 288 single cells per day, this method transforms the scale at which biological and clinical questions can be addressed, opening the door to proteome-wide investigations in single-cell atlases and translational studies.

To date, the authors have analysed more than 4,000 samples using the described platform, and continued advances in mass spectrometry performance mean that a throughput of 500–1,000 cells per day at a comparable depth appears to be within reach.

Overall, it is a very important manuscript that almost eliminates the disadvantages of “label free SCP” and is therefore very important for the field. I therefore recommend a minor revision of this manuscript.

Major issues:
I suspect that most data have been searched against a library or with the MBR option.
I'm more “old school” here and would like to see data where every single cell or 250 pg standard is shown without a match between runs—in other words, unadulterated.
The reason for this is that without a match between runs, outliers that could be instrumental in origin could be seen much earlier.
If possible, I would like to see figures for 1000 injections of single cells and 250 pg HeLa in succession to determine how stable the system is in practice (x-axis number of injections, y-axis protein number without MBR).

Minor items:
Line 22: What does it mean “nearly” 288 samples per day. Please be more accurate.
Would be if interest to see the CV from the 2 parallel columns (rt, quant, ionization efficiency,…). Any kind of data are welcome.

In my opinion, adding videos is a very good idea. They help me understand the workflow much better.
Figure 4: at my personal opinion its not a good idea to show any kind of CV from single cells. In single-cell proteomics, we want to show the difference between individual cells—clear separation is important here.

Figure 5: Please use different blots here.
Please use our publication – figure 3 (Bubis et al. Nat. Met.) as an example. Perhaps it is even possible to compare the accuracy and precision of quantification with our data. Or additional with the data published by the Olsen lab (Astral – not Astral zoom)?

Inflammatory stimulation experiment: In my opinion, it is not enough to compare up-regulated and down-regulated proteins overall. It would be nice if we could use the power of single-cell proteomics to also see differences in the same condition, if

possible. Any kind of deeper evulation is welcome.

Data analysis and bioinformatics: More search parameters should be specified here. In particular, whether the search is performed with or without MBR. Whether a library was searched and what effect this has.

Finally, I would like to congratulate the authors on this excellent manuscript and hope that my suggestions will help to improve its quality.

Reviewer #2

(Remarks to the Author)
Wang et al., describe an excellent workflow with the capacity to revolutionize label free (LF) single cell proteomics. One of the main current limitations of field is the reduced sample throughput, with most recent studies selecting using between 40-80 samples per day. This makes large studies with thousands of cells expensive and time consuming. The workflow described here has the potential to make LF single cell proteomics scale to the tens if not hundreds of thousands of cells. A 5 min gradient identifying >3,000 proteins with comparable coefficients of variation to gradients that are 5x slower is impressive. This work is likely to have big impact in the field.
However, the biological use cases selected to show the value of the workflow have some issues. The HeLa and K562 work compared a cervical cancer cell line to a lymphoblast cell line, hence very pronounced differences between them are expected. As such, it is not a very effective use case to determine the capabilities of the workflow. Furthermore, the use case which could test its biological capacities, the RAW264.7 cells treated with with LPS, shows no real separation between the control and LPS treatment, which is problematic.
It is this Reviewers opinion that the workflow is excellent and of great value to the community, but the use cases need to be improved, in particular the issues with the RAW264.7. There is a good chance the strange results might be more about culture conditions and stimulation more than MS data quality, so it would be beneficial to have this redone. Normally requiring more single cell proteomics experiments would be unreasonable, however with the 288 sample a day workflow, some of the requested changed would require less than 1 day of instrument time.

Main issues:
1- The use cases are not well suited to highlight the capacities of the workflow. Differentiating HeLa cells and K592 cells is not a benchmark, it is more of a common-sense bare minimum. I.E. if it couldn't differentiate these two cell types it would highlight serious concerns.
The LPS study could evaluate the capacities, however the data does not find clear separation between LPS treated RAW264.7 and the control RAW264.7. LPS is quite a strong stimulus, yet the data shows no real separation between control and LPS treatment in component 1 which explains most of the variance, and little separation on component 2. The study which the authors cite (Woo et al.) does show clear separation of LPS and controls.
It is possible the LPS dose in insufficient, or that adding the 400 µL PBS to the control cells caused issues, in either case it is problematic to see no separation. It would be important for the authors to replicate the culture and stimulus conditions used by Woo et al. to show that their method also has the capacity to differentiate stimulated and unstimulated cells. At the moment their RAW264.7 data raises more questions instead of providing a use case to show what the workflow can achieve.

2- Coefficients of variation are used heavily throughout the paper however there are no details on how they are calculated. There is no information to see if normalisation was performed before the CV calculation, or if the CVs were calculated on the log transformed data. Furthermore, there are Spectronaut parameters that are turned on by default, such as the top 3, that heavily affect CVs reducing them considerably, this needs to be at least mentioned so readers can interpret the data accordingly. Ideally to test the instrument performance with their new workflow, the CVs would be calculated without top 3 and on the raw (non-normalised) intensity to really map instrument variability. Guidelines for CVs are available and can be followed.
It also appears that any protein > 1.5 IQR was removed, so unlike other studies which calculate CVs on all proteins identified, this is calculated excluding the most variable proteins. Please at least provide the full plots with these proteins included in the supplemental data.
3- The bar plots used (Fig. 2A, 3A..), are very not suitable for the single cell data. The bar plots with error bars don't show the underlying distribution of the data points. This is an easy point for the authors to address, please just replace them with box plots if the authors don't want to repeat the violin plots from the figures next to them.
4- There are no carry over tests at the single cell level. It would also be insightful to show the chromatograms with blanks as well as the master mix only runs, as currently the carryover was only evaluated with three blanks following a 2 ng HeLa digest on an exploris 480 and not at single-cell input levels or across extended operation on the Orbitrap Astral

Specific points:

1. Please show the data from Figure 2B also as a box plot. I understand the graphical purpose and visuals of 2B but it doesn't help to quantitatively evaluate drift, so in addition to it, having a box plot would be important.
2. For all the figures showing the CVs if the axis has been cut for visualisation purposes and the outliers are not being shown, which is likely the case and can be reasonable to do, please specify it on the figure legend.
3. 250pg of HeLa digest are not very comparable to an average single cell. It could be argued it is not even comparable to an average cancer cell line. In general, at best the protein content without any loses of a HeLa is 250pg. After processing and

digestion, the actual content injected is likely to be significantly less. Furthermore, a lysate has different concentration of proteins than a single cell, so again a lysate would require even lower to be equivalent to a single cell. So, in this case it is not a commonly accepted proxy for a single cell. Minor issue but please correct it in the text.
4. The methods provide no details for the Gene Ontology analysis. This should be fixed and the details of the tool and procedure should be specified. Based on the number of proteins significantly changed which were used for the analysis, it looks reasonably likely the authors did not change the background of this analysis to the proteins detected in the study. This can produce false positives enrichment results. Please set the right background (proteins used for the volcano plot) before doing the GO analysis.
5. Minor comment: The DAVID website is unavailable and the it's previous last full update was in 2021, so it might be helpful to use a more up to date tool such as https://www.webgestalt.org/
6. No information on how the adjusted p-values were calculated is provided within the methods section. For single cell data is recommended to use the more stringent Bonferroni test, however no information is provided so it can't be evaluated. Please specificy this in the methods and confirm it uses a stringent test.
7. There are no details about the data analysis procedure for the HYE work. It is assumed it is done on Spectronaut, but making it more clear would be helpful to readers. Furthermore, what type of normalisation was used for the analysis? Was it normalised for each species? Details might be available on the Navarro paper, but they should be specified in the methods section of this manuscript as well to confirm what was done and to help readers from a broader audience.
8. The HYE data looks good, and it shows the capacity to produce accurate data, however this calculated by removing any protein with a fold change greater or smaller than 1.5 IQR, this could represent the 25% most variable proteins (based on the number of proteins identified). What does the data look like when the IQR filtering is not applied? It appears this analysis has been done with Spectronaut, which is of course perfectly fine, however DIA-NN excels at this type of analysis, would be very insightful to also run the search and analysis on DIA-NN.
9. The HYE data appears to have reasonable accuracy, but much lower precision as there is a very significant spread shown on the box plots. Can the authors comment on why this is?
10. The manuscript is heavy on the technical side, which is fine for proteomics readers, but for a broader audience could use more explanations that simplify the concepts a bit.

In summary, the technical developments are excellent, and this is going to have great impact on the field. The authors just need to work on their data analysis and on the biological use cases which exemplify the potential of this workflow.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I would like to extend my sincere thanks to the authors for their detailed and excellent answers to all of my questions. Thanks to the thorough revision of the manuscript, it has been significantly improved and is now also of great interest to a broad readership.

Through a large-scale study, the authors now demonstrate the method's enormous throughput. In my opinion, the expanded RAW2674.7 dataset is fantastic and addresses very interesting biological questions.

I have no further questions regarding this manuscript.

In closing, I would like to take this opportunity to wish the authors all the best with their manuscript (KM).

Reviewer #2

(Remarks to the Author)
The implemented changes to the manuscript now resolve the main issues I had described, and show that the 288 SPD methods can resolve biological states in cells that are a bit more challenging and interesting that the standard cancer cell lines. Specifically, the new extended data for HeLa single cells show stability and remarkable consistency across the 1000 runs and the RAW264.7 show the difference in inflammatory states upon LPS treatment. The new figures represent the data well. I am happy to accept the manuscript with some small minor corrections assuming the major problem with the PRIDE submission is sorted, but as I mention in the text below I think that is likely a minor error.

• Major issue (but likely a small mistake): PXD080394 currently only has 21 files in total. And few raw files and without large zip files containing collections of raw files. And this is the only accession reported for the paper. I assume something went wrong and all the other raw files are in the previous accession PXD070201, but it is not listed here. This needs to be corrected before it is accepted. I will leave it to the editorial team and authors to sort this minor issue as I don't doubt the authors intent to share all raw files.
• Figure 2: add significance to the boxplots in C. The data looks pretty identical so it's like only adding an N.S label, nothing big.
• Figure 3B: After filtering are all proteins with a CV under 50% or is the graph being limited to 50% for visualisation. It appears to be limited, which is fine, but state it on the legend.

• Figure 4: 14% single cell IDs with much lower Identifications is quite acceptable, some cells will be smaller dying or dead. I understand why its removed, but it would be helpful to have all cells in a supplemental figure.
• Figure 4 it would be good to write in the main text how these data were searched and how the comparison data were searched. I agree it is comparable but some of those datsets are searched with more stringent parameters.
• Don't have line number in this version but it would be very helpful to establish what the long and short term analyses for figure 5 are at the start of the section. The describe the long one and then the short one as is now. The figure appears first without explanation of what the short and long-term references are which is a bit confusing to read.
• Figure 5B&C: please turn these panels into either box plots or density plots to see the overall distribution of identifications. Bar plots without error bars are not appropriate. I suspect the data from Bubis is tighter for IDs and absolute error (compared to the extended) but that is fine, this higher throughput. It would just be good to be able to see that.
• Figure 5D&E: it is very interesting that the extended acquisition seems to have much higher precision than the short term reference, can they author hypothesise why this is?
• Figure 6: Impressive scale to analyse 1,000 cells for LPS treatment
• Is the size difference of the cells between PBS and LPS significant? I think its an interesting result to include in a figure.
• Figure 7: GO analysis without a p-value filtering isn't appropriate here. There needs to be a defined foreground for a GO analysis, without the significance filter setting up foreground and background is complex. But a GSEA providing all the list would fit well for this approach. The authors are correct to avoid double dipping, by defining clusters and then using the defined clusters for differential expression, but just conver the GO section to GSEA. The analysis is quick I'm sure the results will be similar and it will be more robust.
• How were missing values handled for the UMAPs, its not vital but it is helpful to know and would be good to specify in the methods.
• Minor but important issue: Clusters should not be done based on UMAP projections as distances in UMAPS do not have concrete meaning. Ideally the clustering should be done separately and the clusters shown on the UMAP. As this is not a heavily biological paper the matter is less important, but it would be ideal to fix this analysis detail. This paper might be used to check how to analyse single cells, so its to use best practices in the analysis so errors are not replicated.
• For the UMAPs in Figure 7 it would be very helpful to keep the cluster labels at the right locations.
• Figure 7 has plenty of panels available, so why not add more proteins, say Stat3 and NCF2 as extra UMAPs covering inflammatory and antimicrobial.

Summary of Major Revisions

We sincerely thank both reviewers for their careful evaluation of our manuscript and for their constructive suggestions. We were encouraged that both reviewers recognized the potential impact of the 5-minute multicolumn LC-MS workflow for increasing the throughput of label-free single-cell proteomics. We also appreciate that the reviewers' comments helped us identify key areas where the manuscript could be substantially strengthened. These comments led us to expand both the technical and biological evaluation of the platform.

We believe these revisions and the added data significantly strengthen the manuscript and broaden its impact. Whereas the original submission primarily demonstrated the technical capability of the multicolumn LC-MS platform, the revised manuscript now also demonstrates the type of large-scale biological analysis enabled by this throughput. In particular, the expanded RAW264.7 dataset shows that the platform can resolve heterogeneous single-cell proteomic responses within a defined inflammatory perturbation, while the longitudinal HeLa single-cell and QC dataset supports the robustness of the system during extended operation. Together, these additions more clearly demonstrate how the 288-SPD workflow can enable large-cohort label-free SCP studies while maintaining sufficient depth and quantitative consistency for biological interpretation.

Response to Reviewer #1

In this fascinating and scientifically interesting manuscript, the authors describe a high-throughput multicolumn LC system that is a valuable addition to the field of SCP, helping to evolve its throughput to the next level and transition the field from method development to the biological toolbox. By enabling the routine analysis of 288 single cells per day, this method transforms the scale at which biological and clinical questions can be addressed, opening the door to proteome-wide investigations in single-cell atlases and translational studies.

To date, the authors have analyzed more than 4,000 samples using the described platform, and continued advances in mass spectrometry performance mean that a throughput of 500–1,000 cells per day at a comparable depth appears to be within reach.

Overall, it is a very important manuscript that almost eliminates the disadvantages of "label free SCP" and is therefore very important for the field. I therefore recommend a minor revision of this manuscript.

We sincerely thank Reviewer 1 for speaking highly of our study and for recognizing its potential impact on the field.

I suspect that most data have been searched against a library or with the MBR option. I'm more "old school" here and would like to see data where every single cell or 250 pg standard is shown without a match between runs—in other words, unadulterated. The reason for this is that without a match between runs, outliers that could be instrumental in origin could be seen much earlier.

We thank the reviewer for this important suggestion. We agree that evaluating single-cell and low-input standard data without match-between-runs provides a more direct view of the raw performance of the LC-MS platform and can make potential instrumental outliers easier to identify. In response, all datasets used to evaluate system performance were searched using DIA-NN 2.2.0 Academia without match-between-runs. These analyses include the column-to-column reproducibility analysis **(Figure 3),** the longitudinal stability analysis containing 1,230 isolated HeLa single cells and 138 interspersed 250 pg HeLa digest QC injections (**Figure 4),** and the expanded HYE quantitative benchmarking analysis **(Figure 5)**. We have also revised the Data Analysis and Bioinformatics section to clarify the search strategy and explicitly state search settings.

If possible, I would like to see figures for 1000 injections of single cells and 250 pg HeLa in succession to determine how stable the system is in practice (x-axis number of injections, y-axis protein number without MBR).

We thank the reviewer for this helpful suggestion. We have added a new longitudinal stability analysis to directly evaluate system performance during extended single-cell acquisition. As requested, the revised **Figure 4A** plots each injection as a function of acquisition order, with the x-axis showing the number of injections and the y-axis showing the number of protein groups identified without match-between-runs. This analysis includes 1,230 isolated single HeLa cells and 138 interspersed 250 pg HeLa digest QC injections acquired across four sample plates over six days. The 250 pg QC injections remained stable throughout the acquisition sequence, while the single-cell runs showed broader variation, but no systematic drift over extended system operation. The corresponding distributions are summarized in **Figure 4B**. These results demonstrate that the system remained stable during extended operation and that potential outliers can be evaluated directly at the individual-run level.

Line 22: What does it mean "nearly" 288 samples per day. Please be more accurate.
To clarify, the multicolumn workflow can generate one chromatogram every 5.0 minutes. However, Xcalibur introduces a short overhead between consecutive runs during the transition from the previous acquisition to the next acquisition, including method downloading and instrument-control response. In our current setup, this overhead is 20 s between runs. This is why we describe the throughput as "nearly" 288 samples per day rather than exactly 288 samples per day, and this is clarified in the last sentence of Multicolumn nanoLC Setup section in Method.

Would be if interest to see the CV from the 2 parallel columns (rt, quant, ionization efficiency,…). Any kind of data are welcome.
We have added a box plot to **Figure 2C** showing the retention-time distributions from the two analytical columns for four peptides across 1,200 HYE standard runs. We also added a new column-to-column reproducibility analysis in the revised **Figure 3**. Using 250 pg HeLa digest standards, the two analytical columns showed highly similar proteome coverage and CVs. Together, these analyses provide direct RT and quantitative readouts of the variation between the two parallel columns.

In my opinion, adding videos is a very good idea. They help me understand the workflow much better.
We thank the reviewer for this positive comment.

Figure 4: *at my personal opinion its not a good idea to show any kind of CV from single cells. In single-cell proteomics, we want to show the difference between individual cells—clear separation is important here.*
We agree with the reviewer that CV across a heterogeneous single-cell population is not the most appropriate metric for evaluating system performance. In response, we revised **Figure 4** to focus on longitudinal proteome coverage rather than single-cell CV. Separately, we retained CV analysis for the RAW264.7 dataset only at the cluster level, where it is used to describe within-cluster variation among biologically defined cell populations rather than system performance.

Figure 5: *Please use different blots here. Please use our publication – figure 3 (Bubis et al. Nat. Met.) as an example. Perhaps it is even possible to compare the accuracy and precision of quantification with our data. Or additional with the data published by the Olsen lab (Astral – not Astral zoom)?*
We revised **Figure 5** following the plotting style suggested by the reviewer, using Figure 3 from Bubis et al. as a guide. The revised **Figure 5D–F** now includes local CV distributions, protein-level log2 fold-change versus abundance plots, and KDE distributions of measured fold changes, allowing both quantitative precision and accuracy to be evaluated.
We also added a comparison with a low-input Astral HYE dataset from Bubis et al. Raw files from this dataset were downloaded and processed using the same DIA-NN and post-search data processing workflow used for our data, and the comparison is now shown together with our extended-acquisition and short-term reference HYE datasets in **Figure 5**. This comparison showed that our 288-SPD workflow had modestly lower proteome coverage but maintained competitive fold-change accuracy. In particular, the short-term reference dataset showed very low absolute error, while the extended-acquisition dataset maintained comparable quantitative accuracy across more than 1,100 LC-MS runs acquired over four days.

We did not include the Olsen lab dataset because the available HYE comparison was acquired at a substantially higher input amount and therefore was not directly comparable to the low-input HYE analysis in this study.

Inflammatory stimulation experiment: In my opinion, it is not enough to compare up-regulated and down-regulated proteins overall. It would be nice if we could use the power of single-cell proteomics to also see differences in the same condition, if possible. Any kind of deeper evulation is welcome.
We agree. In response, we repeated the RAW264.7 LPS stimulation experiment and expanded the analysis to 974 retained single cells, including 211 PBS-treated cells and 783 LPS-treated cells. In the revised manuscript, UMAP-based clustering showed that the LPS-treated cells separated into two distinct populations **(Figure 6A).** Technical variables, including sorting gate, analytical column, and acquisition order, were not the primary drivers of this separation **(Figure 6B–D).** We further characterized the PBS, LPS-1, and LPS-2 populations using cluster-level proteome coverage, protein-intensity variation, Venn analysis, fold-change distributions, GO enrichment, and representative protein-intensity visualization **(Figures 6E–G and 7A–E).** These analyses revealed an immune-activated LPS population and a second stress-associated LPS population, demonstrating that the platform can resolve heterogeneous responses within the same treatment condition.

Data analysis and bioinformatics: More search parameters should be specified here. In particular, whether the search is performed with or without MBR. Whether a library was searched and what effect this has.
We revised the Methods, "Data Analysis and Bioinformatics" section to more clearly describe the updated search workflow. We also explicitly clarified the use of experimental spectral libraries, match-between-runs, and 10 ng standard/reference runs. For the system-performance analyses **(Figure 3-5),** searches were performed without MBR and without inclusion of the 10 ng standard runs to preserve direct assessment of run-to-run system variation, including the column-to-column comparison, longitudinal single-cell HeLa analysis and HYE mixed-species standards. In contrast, the RAW264.7 biological dataset **(Figures 6–7**) was searched with MBR and 10 ng bulk-digested RAW264.7 reference runs to maximize proteome coverage for downstream biological interpretation.

Respond to Reviewer #2

Wang et al., describe an excellent workflow with the capacity to revolutionize label free (LF) single cell proteomics. One of the main current limitations of field is the reduced sample throughput, with most recent studies selecting using between 40-80 samples per day. This makes large studies with thousands of cells expensive and time consuming. The workflow described here has the potential to make LF single cell proteomics scale to the tens if not hundreds of thousands of cells. A 5 min gradient identifying >3,000 proteins with comparable coefficients of variation to gradients that are 5x slower is impressive. This work is likely to have big impact in the field.

However, the biological use cases selected to show the value of the workflow have some issues. The HeLa and K562 work compared a cervical cancer cell line to a lymphoblast cell line, hence very pronounced differences between them are expected. As such, it is not a very effective use case to determine the capabilities of the workflow. Furthermore, the use case which could test its biological capacities, the RAW264.7 cells treated with with LPS, shows no real separation between the control and LPS treatment, which is problematic.

It is this Reviewers opinion that the workflow is excellent and of great value to the community, but the use cases need to be improved, in particular the issues with the RAW264.7. There is a good chance the strange results might be more about culture conditions and stimulation more than MS data quality, so it would be beneficial to have this redone. Normally requiring more single cell proteomics experiments would be unreasonable, however with the 288 sample a day workflow, some of the requested changed would require less than 1 day of instrument time.

We sincerely thank Reviewer 2 for the careful evaluation of our manuscript and for recognizing the potential impact of the workflow for scaling label-free single-cell proteomics. We also appreciate the reviewer's thoughtful criticism of the biological use cases in the original submission. This comment substantially shaped the revision. In response, we removed the HeLa/K562 biological use case, repeated the RAW264.7 LPS stimulation experiment, and expanded the analysis to better demonstrate single-cell heterogeneity within a defined inflammatory perturbation. We believe these changes significantly strengthened the biological impact of the manuscript. Detailed responses to each point are provided below.

Main issue 1.1-*The use cases are not well suited to highlight the capacities of the workflow. Differentiating HeLa cells and K592 cells is not a benchmark, it is more of a common-sense bare minimum. I.E. if it couldn't differentiate these two cell types it would highlight serious concerns.*

We agree with the reviewer that comparing two distinct cell lines (HeLa/K562) was not the strongest biological use case for demonstrating the capability of the workflow. In response, we removed the HeLa/K562 biological comparison from the revised manuscript and replaced it with a substantially expanded RAW264.7 LPS stimulation experiment, which is detailed in the response to Main issue 1.2.

The LPS study could evaluate the capacities, however the data does not find clear separation between LPS treated RAW264.7 and the control RAW264.7. LPS is quite a strong stimulus, yet the data shows no real separation between control and LPS treatment in component 1 which explains most of the variance, and little separation on component 2. The study which the authors cite (Woo et al.) does show clear separation of LPS and controls.

It is possible the LPS dose in insufficient, or that adding the 400 µL PBS to the control cells caused issues,

in either case it is problematic to see no separation. It would be important for the authors to ***replicate the culture and stimulus conditions used by Woo et al****. to show that their method also has the capacity to differentiate stimulated and unstimulated cells. At the moment their RAW264.7 data raises more questions instead of providing a use case to show what the workflow can achieve.*

We agree with the reviewer that the original RAW264.7 LPS experiment did not provide a sufficiently strong biological use case for demonstrating the capability of the workflow. In response, we repeated the entire LPS stimulation experiment using 100 ng/mL LPS for 24 h, with a matched PBS vehicle control. The revised experiment includes 974 retained single RAW264.7 cells, including 211 PBS-treated cells and 783 LPS-treated cells.

In the revised manuscript, UMAP analysis now shows clear separation between the PBS-treated cells and LPS-treated cells, with the LPS-treated cells further separating into two distinct populations **(Figure 6A).** We also evaluated technical variables, including sorting gate, analytical column, and acquisition order, and these variables did not explain the observed clustering pattern **(Figure 6B–D**). We then performed additional cluster-level analyses, including proteome coverage, protein-intensity variation, fold-change distributions, GO enrichment, and representative protein-intensity visualization **(Figures 6E–G and 7A–E)**. These analyses show that the revised RAW264.7 dataset provides a stronger biological demonstration of the workflow and can resolve heterogeneous single-cell responses within the LPS-treated condition.

Main issue 2.1-*Coefficients of variation are used heavily throughout the paper however there are no details on how they are calculated. There is no information to see if normalisation was performed before the CV calculation, or if the CVs were calculated on the log transformed data*

We revised the Methods, "Data Analysis and Bioinformatics" section to more clearly describe how CV values were calculated and how normalization was applied for different datasets. In general, proteins with more than 33% missing values within the relevant experimental group were excluded, and CVs were calculated from raw protein-level intensities after median normalization within the corresponding group. CVs were calculated from raw intensity values, not from log-transformed values.

Because the datasets served different purposes, the normalization strategy was specified separately:

- For the 250 pg HeLa digest column-to-column and sampling analyses, intensities were median-normalized within the corresponding replicate group before CV calculation.
- For the HYE benchmark, intensities were median-normalized within each HYE-A or HYE-B group across all three species together, which corrects run-level variation while preserving the expected abundance differences between the two mixtures.
- For the RAW264.7 dataset, CVs were calculated within each biologically defined cluster to describe within-cluster variation; separate log2 transformation, normalization, and imputation steps were used for UMAP and fold-change analyses rather than for reporting CVs. These details are now described in the revised Methods section.

Main issue 2.2-*Furthermore, there are Spectronaut parameters that are turned on by default, such as the top 3, that heavily affect CVs reducing them considerably, this needs to be at least mentioned so readers can interpret the data accordingly. Ideally to test the instrument performance with their new workflow, the CVs would be calculated without top 3 and on the raw (non-normalised) intensity to really map instrument variability. Guidelines for CVs are available and can be followed.*

We agree with the reviewer that protein-intensity CVs can be strongly affected by software-specific quantification settings and data-processing choices, and that these details should be clear for readers. In the revised manuscript, all datasets were reprocessed using DIA-NN rather than Spectronaut. Therefore, the Spectronaut-specific "top 3" protein quantification setting is no longer applicable to the revised analysis.
For the system-performance datasets, we used the protein output (pg.matrix) from DIA-NN and clarified the CV calculation and normalization procedure in the revised Methods, "Data Analysis and Bioinformatics" section. We also avoided relying only on CV as the primary readout of instrument performance. In response to Reviewer 1 and Reviewer 2's concerns, we added the longitudinal protein-identification analysis without match-between-runs **(Figure 4),** which allows each single-cell and QC injection to be evaluated directly as a function of acquisition order. We believe this provides a more transparent view of system stability and potential run-level outliers than reporting CV alone.

Main issue 2.3-*It also appears that any protein > 1.5 IQR was removed, so unlike other studies which calculate CVs on all proteins identified, this is calculated excluding the most variable proteins. Please at least provide the full plots with these proteins included in the supplemental data*.
We agree that filtering high-CV proteins can affect the visualized CV distributions and should be reported transparently. In the revised manuscript, the main CV plots still use the 1.5× IQR filter to reduce the influence of extreme outliers and improve visualization of the central distribution. However, to address the reviewer's concern, we now provide the corresponding pre-filter and post-filter statistics in the **Table S1**. Across all datasets, greater than 95% of proteins were retained after IQR filtering, indicating that the filter had minimal impact on protein selection. We also clarified in the Methods where the 1.5× IQR filter was applied.

Main issue 3-*The bar plots used (Fig. 2A, 3A..), are very not suitable for the single cell data. The bar plots with error bars don't show the underlying distribution of the data points. This is an easy point for the authors to address, please just replace them with box plots if the authors don't want to repeat the violin plots from the figures next to them*.
We agree with the reviewer and have replaced the relevant bar plots with box plots in the revised figures.

Main issue 4-*There are no carry over tests at the single cell level. It would also be insightful to show the chromatograms with blanks as well as the master mix only runs, as currently the carryover was only evaluated with three blanks following a 2 ng HeLa digest on an exploris 480 and not at single-cell input levels or across extended operation on the Orbitrap Astral*
We agree that carryover should be evaluated under low-input conditions relevant to single-cell proteomics and on the same MS platform used for the revised study. In response, we revised the carryover experiment using 250 pg HYE samples analyzed on the Orbitrap Astral, rather than the previous 2 ng HeLa digest experiment on the Orbitrap Exploris 480. The revised **Figure 2B** now shows the chromatogram of the 250 pg HYE sample together with the corresponding carryover run. The average carryover intensity was 0.61% relative to the sample signal.

Minor Issue 1-*Please show the data from Figure 2B also as a box plot. I understand the graphical purpose and visuals of 2B but it doesn't help to quantitatively evaluate drift, so in addition to it, having a box plot would be important.*

We added the requested box-plot visualization to the revised **Figure 2C**. The updated panel now shows both the retention-time trajectories across 1,200 continuous HYE standard runs and the corresponding box plots for the four representative peptides, allowing the retention-time variation between the two analytical columns to be evaluated more quantitatively.

Minor Issue 2-*For all the figures showing the CVs if the axis has been cut for visualisation purposes and the outliers are not being shown, which is likely the case and can be reasonable to do, please specify it on the figure legend*

We agree that the figure legends should clearly indicate how CV distributions are displayed. In the revised CV figures, the CV axes were not truncated for visualization. The main CV plots use the 1.5× IQR-filtered dataset, as described in our response to Main issue 2.3, and the corresponding pre-filter and post-filter statistics are provided in **Table S1**. We have also clarified in the figure legends where the 1.5× IQR filter was applied.

Minor Issue 3-*250pg of HeLa digest are not very comparable to an average single cell. It could be argued it is not even comparable to an average cancer cell line. In general, at best the protein content without any loses of a HeLa is 250pg. After processing and digestion, the actual content injected is likely to be significantly less. Furthermore, a lysate has different concentration of proteins than a single cell, so again a lysate would require even lower to be equivalent to a single cell. So, in this case it is not a commonly accepted proxy for a single cell. Minor issue but please correct it in the text.*

We revised the wording throughout the manuscript from "single-cell equivalent" to "single-cell-level" to avoid implying that the 250 pg digest standard is a direct substitute for a processed single-cell sample. We also added citations to recent SCP method-development studies that used 250pg HeLa digest as a technical benchmark for low-input system performance.

Minor Issue 4-*The methods provide no details for the Gene Ontology analysis. This should be fixed and the details of the tool and procedure should be specified. Based on the number of proteins significantly changed which were used for the analysis, it looks reasonably likely the authors did not change the background of this analysis to the proteins detected in the study. This can produce false positives enrichment results. Please set the right background (proteins used for the volcano plot) before doing the GO analysis*

We have revised the Methods section to provide additional details regarding the Gene Ontology enrichment analysis. Specifically, we clarified that GO enrichment was performed using upregulated or downregulated proteins identified from the fold change analysis as the input list, with *all* proteins included in the corresponding fold change analysis used as the background population. Therefore, the enrichment analysis was performed against the experimentally detected protein set rather than the full proteome, minimizing the potential for false-positive enrichment results. The revised Methods section now explicitly describes the enrichment tool, parameters, input lists, and background selection. The complete fold-change results, DAVID GO-term outputs, assigned biological-theme categories, and the supporting source used for each GO-term category assignment are provided in **Table S2**

Minor Issue 5-*The DAVID website is unavailable and the it's previous last full update was in 2021, so it might be helpful to use a more up to date tool such as https://www.webgestalt.org/*

We appreciate the reviewer's suggestion to consider a more recently updated enrichment-analysis platform. During the revision, DAVID was accessible, and according to the DAVID documentation, DAVID 2025 was released in March 2026 with updated documentation and knowledgebase content. To address the reviewer's concern, we evaluated the enrichment results using both DAVID and WebGestalt. The two platforms produced comparable biological interpretations and similar enrichment patterns. WebGestalt returned a larger number of enriched GO terms than DAVID, but these additional terms were largely overlapping or redundant and did not change the primary biological conclusions.

We therefore retained the DAVID-based analysis in the revised manuscript because its functional clustering output provided a more concise and readily interpretable summary of enriched biological processes and facilitated manual verification of the GO-term clusters reported in the study. Given that WebGestalt produced consistent biological conclusions in our comparison, we will also consider using it in future studies where its broader GO-term output may provide additional value.

Minor Issue 6-*No information on how the adjusted p-values were calculated is provided within the methods section. For single cell data is recommended to use the more stringent Bonferroni test, however no information is provided so it can't be evaluated. Please specificy this in the methods and confirm it uses a stringent test.*

We agree that the method for calculating adjusted p-values should be clearly described when adjusted p-values are used. In the revised manuscript, we no longer report adjusted p-values to support the biological conclusions.

Instead, the RAW264.7 analysis is now presented as an exploratory cluster-level analysis of single-cell heterogeneity. To avoid over-interpreting differences among clusters defined from the same dataset, we focus on fold-change patterns, GO enrichment, and representative protein-intensity trends rather than formal differential-abundance significance testing. Accordingly, adjusted p-value calculations have been removed from the revised manuscript and the Methods section.

Minor Issue 7-*There are no details about the data analysis procedure for the HYE work. It is assumed it is done on Spectronaut, but making it more clear would be helpful to readers. Furthermore, what type of normalisation was used for the analysis? Was it normalised for each species? Details might be available on the Navarro paper, but they should be specified in the methods section of this manuscript as well to confirm what was done and to help readers from a broader audience*

In the revised manuscript, we added HYE-specific details to the Data Analysis and Bioinformatics section. Briefly, the newly acquired HYE datasets were processed using DIA-NN 2.2.0 Academia rather than Spectronaut, and the analyses were performed without match-between-runs and without inclusion of the 10 ng standard runs.

We also clarified the normalization procedure used for the HYE analysis. Protein intensities were median-normalized within each HYE-A or HYE-B group across all three species together, rather than normalized separately for each species. This normalization strategy reduces run-level technical variation while preserving the expected species-specific abundance differences between HYE-A and HYE-B for fold-change evaluation. These details have now been specified in the revised Methods section.

Minor Issue 8-*The HYE data looks good, and it shows the capacity to produce accurate data, however this calculated by removing any protein with a fold change greater or smaller than 1.5 IQR, this could represent the 25% most variable proteins (based on the number of proteins identified). What does the data look like when the IQR filtering is not applied? It appears this analysis has been done with Spectronaut, which is of course perfectly fine, however DIA-NN excels at this type of analysis, would be very insightful to also run the search and analysis on DIA-NN.*

We agree that the effect of the 1.5× IQR filtering should be reported transparently. In the revised manuscript, we provide the corresponding pre-filter and post-filter statistics in **Table S1**. Across the HYE datasets, greater than 95% of proteins were retained after IQR filtering, indicating that the filter removed only a small number of extreme outliers rather than a large fraction of proteins. In addition, the revised HYE datasets **(Figure 5)** were reprocessed using DIA-NN 2.2.0 Academia rather than Spectronaut.

Minor Issue 9-*The HYE data appears to have reasonable accuracy, but much lower precision as there is a very significant spread shown on the box plots. Can the authors comment on why this is?*

We agree that the HYE results **(Figure 5)** require careful interpretation because different metrics capture different aspects of quantitative performance. In the revised Figure 5, run-to-run quantitative precision is evaluated primarily by local CV, whereas accuracy is evaluated by the absolute error between the measured and expected HYE ratios. The fold-change distributions provide an additional visualization of measured ratio behavior, but they are not interchangeable with local CV.

In the extended-acquisition dataset, the measured fold changes remained generally centered near the expected ratios, supporting reasonable quantitative accuracy. However, compared with the short-term reference dataset, the extended-acquisition dataset showed higher local CVs and larger absolute errors, indicating additional run-to-run variation during prolonged acquisition. These results show that local CV, fold-change distribution, and absolute error provide complementary information, and we therefore interpret the HYE benchmark using all three metrics together rather than relying on a single plot.

Minor Issue 10

The manuscript is heavy on the technical side, which is fine for proteomics readers, but for a broader audience could use more explanations that simplify the concepts a bit

We appreciate the reviewer's perspective. In the revised manuscript, we substantially expanded the RAW264.7 LPS stimulation experiment and its accompanying interpretation. This revised section provides a more direct biological example of how increased single-cell throughput can reveal heterogeneous proteomic responses within the same treatment condition. We believe this added use case helps a broader audience better understand the biological value of the workflow.

Reviewer #1 (Remarks to the Author):

I would like to extend my sincere thanks to the authors for their detailed and excellent answers to all of my questions.
Thanks to the thorough revision of the manuscript, it has been significantly improved and is now also of great interest to a broad readership.
Through a large-scale study, the authors now demonstrate the method's enormous throughput. In my opinion, the expanded RAW2674.7 dataset is fantastic and addresses very interesting biological questions.
I have no further questions regarding this manuscript.
In closing, I would like to take this opportunity to wish the authors all the best with their manuscript (KM).

We sincerely thank KM for the positive assessment of the revised manuscript and for the encouraging comments.

Reviewer #2 (Remarks to the Author):

The implemented changes to the manuscript now resolve the main issues I had described, and show that the 288 SPD methods can resolve biological states in cells that are a bit more challenging and interesting that the standard cancer cell lines. Specifically, the new extended data for HeLa single cells show stability and remarkable consistency across the 1000 runs and the RAW264.7 show the difference in inflammatory states upon LPS treatment. The new figures represent the data well. I am happy to accept the manuscript with some small minor corrections assuming the major problem with the PRIDE submission is sorted, but as I mention in the text below I think that is likely a minor error.

• Major issue (but likely a small mistake): PXD080394 currently only has 21 files in total. And few raw files and without large zip files containing collections of raw files. And this is the only accession reported for the paper. I assume something went wrong and all the other raw files are in the previous accession PXD070201, but it is not listed here. This needs to be corrected before it is accepted. I will leave it to the editorial team and authors to sort this minor issue as I don't doubt the authors intent to share all raw files.
Thank you for this comment and for all of your valuable feedback. We have resolved the issue and now all the .raw files and search results have been uploaded with the accession number of PXD081782 in the revised draft.
Project accession: PXD081782
Token: LowMnfyXqRhY

• **Figure 2**: add significance to the boxplots in C. The data looks pretty identical so it's like only adding an N.S label, nothing big.

We evaluated statistical comparisons between the odd- and even-numbered injections. However, each column was represented by ~600 sequential technical measurements. With this large number of observations, a conventional null-hypothesis test is highly sensitive to very small systematic offsets and tests whether the mean retention-time difference is exactly zero, rather than whether the difference is analytically meaningful. Indeed, mean differences of only 0.058 and 0.031 min (3.5 and 1.9 s, respectively) were statistically significant after multiple-testing correction. Therefore, significance labels alone could overstate the practical importance of these small differences. We have instead reported the magnitude of the paired retention-time differences, which more directly describes the agreement between the two analytical columns.

Peptide	Odd mean	Even mean	Raw p value	BH-adjusted p value	Statistical significance
Peptide 1	1.2990	1.3126	0.1035	0.1381	ns
Peptide 2	4.2133	4.1554	3.18×10^{-8}	1.27×10^{-7}	****
Peptide 3	3.2570	3.2258	8.88×10^{-4}	0.00178	**
Peptide 4	2.4010	2.3957	0.5388	0.5388	ns

• **Figure 3B**: After filtering are all proteins with a CV under 50% or is the graph being limited to 50% for visualisation. It appears to be limited, which is fine, but state it on the legend.
The y-axis in Figure 3B is not limited to 50%; the maximum post-filtering CV is 50.5%. The y-axes of all CV plots in the manuscript display the full observed range. We have clarified this in the Figure 3 legend. Detailed pre- and post-filtering CV statistics are provided in Table S2.

• **Figure 4:** 14% single cell IDs with much lower Identifications is quite acceptable, some cells will be smaller dying or dead. I understand why its removed, but it would be helpful to have all cells in a supplemental figure.
We have added a supplemental run-order plot showing all single HeLa cells and QC samples, including those excluded by the 1.5× IQR filter, as Figure S3. The corresponding text in has also been updated to reference this figure.

• **Figure 4** it would be good to write in the main text how these data were searched and how the comparison data were searched. I agree it is comparable but some of those datsets are searched with more stringent parameters.
Because the full set of search settings is extensive and differs across datasets, we provide these details in Table S1 to allow direct comparison without interrupting the flow of the main text. We also added a reference to Table S1 in the main text.

• Don't have line number in this version but it would be very helpful to establish what the long and short term analyses for figure 5 are at the start of the section. The describe the long one and then

the short one as is now. The figure appears first without explanation of what the short and long-term references are which is a bit confusing to read.
We now define the extended-acquisition and short-term reference datasets at the beginning of the section. We also moved Figure 5 to the end of the section so that both datasets are described before the figure is presented.

• **Figure 5B&C:** please turn these panels into either box plots or density plots to see the overall distribution of identifications. Bar plots without error bars are not appropriate. I suspect the data from Bubis is tighter for IDs and absolute error (compared to the extended) but that is fine, this higher throughput. It would just be good to be able to see that.
We agree that the original Figure 5B and 5C bar plots did not clearly distinguish summary values from underlying distributions and have therefore removed these panels. The original intent was to provide concise summaries of the numbers of quantifiable proteins and the deviations between the measured and expected mean fold changes.
In the revised Figure 5A–C, the numbers of quantifiable proteins are labeled directly in the protein-level fold-change scatter plots, and the KDE plots show the complete distributions of measured protein-level $\log_2$ fold changes relative to the expected values. The corresponding mean absolute errors are now reported in the main text. The original HYE-composition panel has been moved to Figure S4A.
In addition, we generated a boxplot showing the run-level distributions of protein identifications for the extended acquisition, short-term reference, and Bubis et al. datasets and included it as Figure S4B and reference it in the main text. As anticipated, the short-term and Bubis et al. datasets show tighter distributions than the extended-acquisition dataset, consistent with their longer acquisition periods and lower-throughput LC method.

• **Figure 5D&E**: it is very interesting that the extended acquisition seems to have much higher precision than the short term reference, can they author hypothesise why this is?
This likely reflects the substantially larger number of replicate measurements used to estimate each protein-level mean fold change in the extended dataset. However, the higher CVs and absolute errors indicate that the short-term reference dataset has better run-to-run reproducibility and quantitative accuracy, as already discussed in the text. We have added a brief clarification to distinguish these measures.

• **Figure 6**: Impressive scale to analyse 1,000 cells for LPS treatment
    • Is the size difference of the cells between PBS and LPS significant? I think its an interesting result to include in a figure.

We agree that relating cell size to the identified clusters would be informative. However, the Tecan Uno did not record the size of individual isolated cells. The average cell diameters for the PBS and LPS treated populations, which were measured before isolation, are reported in the main text.

• **Figure 7:** GO analysis without a p-value filtering isn't appropriate here. There needs to be a defined foreground for a GO analysis, without the significance filter setting up foreground and background is complex. But a GSEA providing all the list would fit well for this approach. The authors are correct to avoid double dipping, by defining clusters and then using the defined clusters for differential expression, but just conver the GO section to GSEA. The analysis is quick I'm sure the results will be similar and it will be more robust.
We replaced the GO analysis in Figure 7 with Gene Set Enrichment Analysis (GSEA) performed using WebGestalt 2024. For each pairwise cluster comparison, proteins were ranked by protein-level log2 fold change, and GSEA was performed using the full ranked protein list. The revised Figure 7A and Table S3 now report the GSEA results, and the main text has been updated to interpret the biological patterns based on this GSEA analysis. The Methods section has also been revised to describe the GSEA workflow.

• How were missing values handled for the UMAPs, its not vital but it is helpful to know and would be good to specify in the methods.
Missing values were imputed using a Missing-Not-At-Random (MNAR) approach with a fixed random seed of 42 after log2 transformation and median normalization across all cells. This information was already included in the Methods section, and we have revised the wording to make the missing-value handling for the UMAP analysis easier to locate.

• Minor but important issue: Clusters should not be done based on UMAP projections as distances in UMAPS do not have concrete meaning. Ideally the clustering should be done separately and the clusters shown on the UMAP. As this is not a heavily biological paper the matter is less important, but it would be ideal to fix this analysis detail. This paper might be used to check how to analyse single cells, so its to use best practices in the analysis so errors are not replicated.
We have revised the workflow and performed Leiden clustering independently of the UMAP projection. The resulting cluster labels are now visualized on the UMAP in Figure 6E.
All downstream cluster-based analyses were updated using the new Leiden cluster assignments, including Figure 6F–H, Figure 7A, and Table S2-3. The Methods section has also been revised to describe the Leiden clustering parameters.

• For the UMAPs in Figure 7 it would be very helpful to keep the cluster labels at the right locations.
The cluster labels (PBS, LPS1, and LPS2) have now been added at their corresponding locations in all UMAP panels in Figure 7.

• Figure 7 has plenty of panels available, so why not add more proteins, say Stat3 and NCF2 as extra UMAPs covering inflammatory and antimicrobial.
UMAPs colored by the protein intensities of STAT3 and NCF2 have now been added to Figure 7. We also added PCNA as a marker of DNA replication and ANXA7 as a marker associated with cytoskeletal organization.