Review COMMONS

JCB

Quantifying organellar ultrastructure in cryo-electron tomography using a surface morphometrics pipeline

Benjamin Barad, Michaela Medina, Daniel Fuentes, R. Wiseman, and Danielle Grotjahn

Corresponding Author(s): Danielle Grotjahn, Scripps Research Institute

Review Timeline:

Submission Date:	2022-04-26
Editorial Decision:	2022-05-08
Revision Received:	2022-09-22
Editorial Decision:	2023-01-03
Revision Received:	2023-01-11

Monitoring Editor: Richard Youle

Scientific Editor: Dan Simon

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

DOI: https://doi.org/10.1083/jcb.202204093

Revision 0

Review #1

1. Evidence, reproducibility and clarity:

Evidence, reproducibility and clarity (Required)

Summary:

Barad and Medina, along with their co-authors, report on the development of a new software toolkit to quantitatively assess membrane structures that are observed in cryo-ET. This new toolkit builds upon existing methodologies by successfully incorporating additional methods and applying this to cryo-ET data to allow for more automated and reliable segmentations. This work addresses a long-standing difficulty in generating membrane segmentations, which are either done manually with huge labor investments or with automated methods that are known to be error prone. The authors demonstrate that their toolkit can generate high quality segmentations across multiple tomograms with limited manual intervention. They use correlative light and electron microscopy in combination with these segmentations to gain insight into the ultrastructural morphology of mitochondria within embryonic fibroblasts, both under control conditions and under endoplasmic reticulum stress induced by treatment with the drug Thapsigarin. Unlike changes to the ER which are more dramatic under stressed conditions, the changes to the mitochondria are more subtle and impossible to quantify without high quality segmentations. The authors show that inner and outer membrane distances change under stress, and that the distances between cristae, their junctions, and the angle of the cristae with respect to the margin of the mitochondria change. While they characterize the curvedness under the same set of conditions, they report no significant differences.

Major comments:

- A major concern is that the data are reported and analyzed on a per tomogram basis when many tomograms contain multiple mitochondria. Given that the mitochondria appear mostly well separated in Sup. Fig 1 with only a few connections visible, and the high degree of pleomorphism noted by the authors, I would strongly suggest that the authors use each mitochondrion as the basis for reporting their metrics rather than the FOV/tomogram as this would avoid mixing metrics from different mitochondria that may be in different states (e.g., fusion/fission). This would apply to data shown in Figures 3, 4, 5, and 6.
- In Figure 3C the authors show the combined distribution of OMM-IMM distances within each condition. This may obscure some variability within populations. Individual histograms for all mitochondria should be included as supplementary material. Currently, it is difficult to judge if the peak of the combined distribution is appropriate and impossible to judge the variability between tomograms (preferably mitochondria, see above comment). Additionally, the shape of the distributions appears significantly different between conditions, suggesting that selecting a single peak value as representative and the basis for the statistical tests (Fig 3D) might not be

appropriate. Please comment.
- In Figure 4C-F, again combined distributions are shown. Authors should include individual histograms for all mitochondria as supplementary material. The diversity of distributions in the metrics are more pronounced than the distances in reported in Fig 3, again making assessment of variability difficult and raising doubt about using the single peak value.
- It would be helpful to include the curvature or curvedness of the OMM for each mitochondrion in the supplementary material. The data to correlate OMM curvature with elongated/fragmented mitochondria should be available and might be of interest to some readers.
- As the work reported here is heavily computational, additional details about the computer hardware used and the time it took for the calculations to complete would be helpful for readers considering applying the code to their own data.
- Discussion should be expanded to include a comparison of semi-automated segmentations generated here versus manual results from Navarro (Ref 35) & Burt (Ref 54 / doi: 10.1371/journal.pbio.3001319) and how one might estimate the error.

Minor comments:

- In the fourth sentence of the third paragraph of the introduction, Hoppe 1992 is cited as evidence of the limitations of work published in 2020, which is confusing. Perhaps the sentence can be re-phrased?
- Pink and purple very close, consider alternative pair of colors or different shades to distinguish OMM and IMM
- For all data, exact n per condition should be given (in text and captions as appropriate), not a range for the whole set.
- Orientation of scaleboxes/scalebars should be consistent per figure panel. If knowledge of the axes is important to the reader, these should be included as well.
- In the last sentence of the introduction, the term "organellar architectures" is used, instead of the previously defined "membrane ultrastructure." Consider changing for clarity.
- Inconsistent use of the phrase "cryo-electron tomography" after defining and using "cryo-ET"
- Authors argue that the distinction between curvedness and curvature is important and that curvature is less appropriate in this context, but then use curvature in the abstract, throughout introduction and in the results section. Usage can be improved for readability.
- In section "Development of a framework to automate quantification of ultrastructural features of cellular membranes" the second last sentence should read "... higher quality membrane surfaces as compared..."
- In section "IMM curvedness is differentially sensitive to Tg treatment in elongated and fragmented mitochondrial networks" the fourth sentence should perhaps read "... despite apparent visual differences, no significant..."
- The term "cell's growth plane" is not clear from the text nor from Fig 6A. Do the authors mean surface of the substrate the cell is growing on?
- In Materials and Methods:
- The authors report that manual back-blotting was used in a Vitrobot. This is non-standard usage and more details should be provided.
- The description of the Leica microscope is insufficient. The objective lens and camera used should be included.
- In section "Fluorescence Guided Milling" in the third sentence, the word "based" is repeated,

second can be removed. A second Pt coat on top of the GIS would also be unusual, please check writing for accuracy.
- Symbol for degree (or the word degree) should be added to angular increment and tilt range for clarity.
- Capitalization of TomoSegMemTV is inconsistent.
- Fig 1B: showing computational steps twice does not provide additional information. Consider just one example. Also, labels for elongated and fragmented would be more useful than the duplicated labels for each computational step.
- Fig 2A caption - should report actual thickness range measured (as given in Materials and Methods section) instead of estimated range.
- Fig 3 title - consider replacing "Inter-mitochondrial membrane..." with "Intra-mitochondrial membrane..." for clarity.
- Fig 3C caption - should explicitly state it is a combined histogram and that the dashed lines correspond to the peak of the pooled data.
- Fig 5E middle, legend obscures some of the data.
- Fig 6B and 6C caption - upper and lower parts not explicitly described.

2. Significance:

Significance (Required)

This work primarily describes a technical advancement in methods to analyze cryo-ET data. The novelty arises from the combination of methods and their application rather than completely new ideas or approaches. Demonstrations of the utility of this toolkit based on the authors' analyses are convincing and will likely help a number of researchers in the field who are engaged in explorations of cellular ultrastructure and organelle responses to stimuli. Importantly, this work will help the field move past qualitative descriptions, historically accepted only because quantitative measurements at this level have not been feasible. Overall, in this reviewer's opinion, while the biological findings are modest, the utility of the toolkit for the field is indisputable and the work is of sufficient quality for publication.

My expertise is in cryo-EM, both single particle analysis and tomography, as well as CLEM workflows, applied mostly to cytoskeletal research and some ER stress. I do not have or strong background in mitochondrial biology nor sufficient computer science expertise to evaluate the numerical methods employed, but based on inspection of the github contents, the screened Poisson reconstruction algorithm is not reimplemented here.

3. How much time do you estimate the authors will need to complete the suggested revisions:

Estimated time to Complete Revisions (Required)

(Decision Recommendation)

Between 1 and 3 months

4. Review Commons values the work of reviewers and encourages them to get credit for their work. Select 'Yes' below to register your reviewing activity at Publons; note that the content of your review will not be visible on Publons.

Reviewer Publons

Yes

Review #2

1. Evidence, reproducibility and clarity:

Evidence, reproducibility and clarity (Required)

Barad, Medina et al. presents a new toolkit for the analysis of membrane ultrastructure in cryo-tomograms. More specifically, the toolkit is designed to compare curvature, angles and spacing between different membrane types in mitochondria. These analyses allow for the quantitative comparison of membrane features e.g. for different growth conditions. To demonstrate the utility of the toolkit tomogram datasets of mitochondria in the presence and absence of ER stress were analyzed. The authors conclude that ER stress affects mitochondria morphology through remodeling of the membrane structure.
The presented biological results and statistics are convincing and show active mitochondrial membrane remodeling in the cell when exposed to ER stress. It is also clear that there is a need for more quantitative evaluation based on the wealth of tomographic image features and mitochondrial membranes are certainly a well-chosen application. For this purpose, the authors developed a new workflow even though most of the discussed analyses are very specific to mitochondrial structures. Therefore, broader applications of these tools to other organelles are not easily envisaged without significant adaption. In that context, the title and abstract overpromise a much more powerful utility that can be applied to any other membrane analysis. Rather it seems that the proposed workflow is more of a specific tool or a pipeline for mitochondrial inner and outer membrane analysis instead of a toolkit for general morphological analysis. Hence, the manuscript cannot be accepted in its current form. In particular, the structure needs a significant rework of editing to become more comprehensible.

Major comments:

1. Title and abstract need to be toned down not to overpromise a very general toolkit. The

presented method may be a tool or a collection of scripts - a toolkit can be used to address other types of (membrane) analysis problems. In the end, the analysis builds to a large extent on the previous developments and implementation of PyCurve. Perhaps, the most interesting contribution here is the application of the mesh generation by the Poisson reconstruction method to the segmented membranes, which is, however, well implemented in the used pymeshlab framework. The computation of distances and angles is straightforward.
2. When reading the manuscript, the reader is left in the open whether this is a method paper or a biological results paper. The title/abstract suggests that this is a method paper and the manuscript is more of a mitochondrial membrane report in ER stress. Therefore, the title/abstract does not reflect the manuscript very well.
3. The manuscript also requires substantial structural editing. Several references to Figures are not appearing in the text in the order that the Figure panels are built. Excessive cross-referencing of figures also make the manuscript hard to read.
4. The focussing to a method paper will also require more in-depth descriptions of the methodology in the main text. Although the code is deposited at github, there is no script-based workflow and description presented in the manuscript. Although Figure 1 puts the work into context of tomography, it remains very superficial on the image analysis. What are the input and output formats required for each step to follow the sequence of the workflow and at which steps critical interactive input is needed? What are the hardware requirements (CPU, GPU) or performance characteristics (CPU hours for certain operations)?
5. Figures 3-7 contain colorful 3D renderings of the measured quantities. In addition, they are filled with histograms of every possible quantitative parameter, which often are not very significant or different between. The authors should focus the main results and the figures to show the most relevant and significant findings and put the remaining panels and results into the supplement.
6. The exact morphological discrimination between fragmented and elongated mitochondria is not easily understood from the results section. What is really meant by blinded manual classification? It only became clear when reading the methods. The results section should stand on its own. How is the overall population between fragmented and elongated cells is affected after Tg application?
7. Similarly, what is meant by manual classification of IMM, OMM and ER? Is there any clustering involved?
8. One of the key steps is the generation of a smooth surface from a segmented membrane, there is a question whether true membrane disruptions will be smoothed and may be overlooked in this approach. When these disruptions present true membrane ruptures, they may be of particular biological importance. The authors should support the choice and selection of the smoothing parameters in order to illustrate this potential pitfall.
9. Throughout the manuscript, the authors mention statistical significance several times and one of the main aims of the study is perform statistical hypothesis testing. It is important to specify the significance test (not only in the methods) and the p-value in order to support this claim. In the manuscript, the authors use exclusively the Mann-Whitney test. What is the rationale for choosing this test? Have the authors considered comparing the total distributions and not just the peaks with e.g. a Kolmogorov-Smirnov test? For a statistical methods paper, there are also no discussion on error analysis.

Minor comments:

1. https://github.com/grotjahnlab/surface_morphometrics should include an example data set or tutorial for dissemination.
2. What is meant by growth plane? This term is not defined in the manuscript.
3. What is meant by vehicle treatment? There is no explanation in the main text of the manuscript.
4. Angle between OMM and cristae: Maybe use the average angle of each cristae for comparison or fit a plane for each cristae because you are interested in the angle between the cristae and the OMM and the membrane of the cristae has a lot of uneven surfaces
5. Have the authors noticed/calculated any differences in the width of the cristae?
6. Methods: Automated surface reconstruction: "In cases where the resulting surface was very complex, the surface was simplified..." How was the complexity determined?
7. Methods: Calculation of distances between individual surfaces: "For surfaces with small numbers of triangles, this was accomplished using a distance matrix...". What is the threshold for a small number of triangles?

2. Significance:

Significance (Required)

The aim of the paper is well motivated. Cryo-ET is a growth field and there is a need for quantitative parameterization of cryo-ET data. Recently a toolkit for the analysis of filaments from cryo-ET has been published (Dimchev et al. 2021 DOI: 10.1016/j.jsb.2021.107808). Given the specific nature of the implementation, i.e. the membrane structures of mitochondria, I cannot easily see that this implementation will be useful beyond the analysis of mitochondrial membrane structure.

3. How much time do you estimate the authors will need to complete the suggested revisions:

Estimated time to Complete Revisions (Required)

(Decision Recommendation)

Between 1 and 3 months

4. Review Commons values the work of reviewers and encourages them to get credit for their work. Select 'Yes' below to register your reviewing activity at Publons; note that the content of your review will not be visible on Publons.

Reviewer Publons

Yes

Revision Plan

Manuscript number: RC-2022-01274
Corresponding author(s): Danielle Grotjahn

1. General Statements [optional]

We would like to thank the reviewers for their prompt and thoughtful input on our manuscript, and their willingness to participate in more portable review through ReviewCommons.

2. Description of the planned revisions

Reviewer #1, major comments:

- A major concern is that the data are reported and analyzed on a per tomogram basis when many tomograms contain multiple mitochondria. Given that the mitochondria appear mostly well separated in Sup. Fig 1 with only a few connections visible, and the high degree of pleomorphism noted by the authors, I would strongly suggest that the authors use each mitochondrion as the basis for reporting their metrics rather than the FOV/tomogram as this would avoid mixing metrics from different mitochondria that may be in different states (e.g., fusion/fission). This would apply to data shown in Figures 3, 4, 5, and 6.

We appreciate the reviewer suggestion to separate on a per mitochondrion vs per tomogram basis for our analysis. While we do not anticipate that this will significantly change the overall findings, we agree that splitting per mitochondrion will account for any possible variability between mitochondria within the given field of view. Furthermore, we anticipate that this will actually improve our analysis and statistical power by effectively increasing the total sample size per experimental group. For our next revision, we will divide surfaces on a per mitochondrion basis within a given tomogram, and re-run the full analysis pipeline. Additionally, per reviewer request, we will include an output histogram for each measurement per mitochondrion surface in a supplemental figure.

- In Figure 3C the authors show the combined distribution of OMM-IMM distances within each condition. This may obscure some variability within populations. Individual histograms for all mitochondria should be included as supplementary material. Currently, it is difficult to judge if the peak of the combined distribution is appropriate and impossible to judge the variability between tomograms (preferably mitochondria, see above comment). Additionally, the shape of the distributions appears significantly different between conditions, suggesting that selecting a single peak value as representative and the basis for the statistical tests (Fig 3D) might not be appropriate. Please comment.

We will include individual histograms for each measurement per mitochondrion surface in a supplemental figure.

We agree that peak-based statistical tests limit our ability to quantify more complex differences, and this is why we chose to output histograms in addition to violin plots, so that shape differences can be observed qualitatively. A major challenge of shape-based statistical quantification is the assessment of independent samples. By using peak-based quantification, we could assume that each tomogram (and in the planned revision, each mitochondrion) is an independent sample, but for shape distribution this is inappropriate since there is more than one value represented per tomogram. Running a KS test with N equal to the number of tomograms yields no significance even in the visible cases where the shape appears very different.

However, the number of triangles also poorly represents the number of independent samples, since 1) the number of triangles used to represent a surface is somewhat arbitrary and remeshing can change it dramatically and 2) Our chosen triangle size is considerably smaller than the visually observed feature size in order to allow effective vector voting in the pycurv AVV algorithm. The result of this is that when we use a KS test on the distribution of values per triangle, even visually identical distributions yield p-values below 10^-200.

We do estimate the approximate smallest feature size during our calculations, since that is used to generate the radius used by pycurv in vector voting, to be 12 nm (the radius hit parameter in pycurv). During a public presentation of this work an audience member suggested that we might use the area implied by this feature size (~450 nm^2) as the size of an independent sample. This would yield around 1000 independent samples per tomogram. Because the choice of feature size is heuristic and manual, this is not as statistically sound as the peak-based metric, which is why we believe that the more conservative peak-based statistical testing is the gold standard for proving differences, but we believe this will be the most reliable way to quantify differences in shape of distributions. We plan to implement this quantification in our revision, and will evaluate whether it gives "expected" statistical results by a bootstrapping approach using subsampling of triangles from the same vs different mitochondria.

We would welcome reviewer suggestions for additional shape-based metrics and will explore other potential metrics to capture shape as part of our revision. While our peak-based metrics demonstrate our ability to statistically capture small changes in ultrastructure with this method, shape-based quantification will significantly enhance the capability to capture finer changes in structure that may be critical to understand physiologically.

Once this additional testing is complete, we will add a section to the results section describing choice of statistical framework. We also plan to generate a supplementary table showing the results of the peak-based quantification alongside all shape-based quantifications.

- In Figure 4C-F, again combined distributions are shown. Authors should include individual histograms for all mitochondria as supplementary material. The diversity of distributions in the metrics are more pronounced than the distances in reported in Fig 3, again making assessment of variability difficult and raising doubt about using the single peak value.
We will include individual histograms for each measurement per mitochondrion surface in a supplemental figure.

As we describe above, we will make test several options for distribution-based statistical quantifications and incorporate the results in the manuscript. We expect them to be useful for every measurement we make.

- It would be helpful to include the curvature or curvedness of the OMM for each mitochondrion in the supplementary material. The data to correlate OMM curvature with elongated/fragmented mitochondria should be available and might be of interest to some readers.
We will calculate curvedness of the OMM for each mitochondrion and include these data in the supplemental material. The inverse of the curvedness of the OMM gives a reasonable approximation of the radius of the mitochondrial "tube", a feature which can be challenging to quantify fully automatically, and we agree that this may be of particular interest to some of our readers – particularly if morphology changes or stress-driven changes alter that radius in a statistically significant way!

Revision Plan

Reviewer #1, minor comments:

- For all data, exact n per condition should be given (in text and captions as appropriate), not a range for the whole set.
We will report the exact n per condition in text and in captions after we separate our data on a per mitochondrion basis and update the analysis.

- Fig 5E middle, legend obscures some of the data.
We will reformat the graph such that the legend does not obscure the data after we separate our data on a per mitochondrion basis and update the analysis.

Reviewer #2, major comments:

Barad, Medina et al. presents a new toolkit for the analysis of membrane ultrastructure in cryo-tomograms. More specifically, the toolkit is designed to compare curvature, angles and spacing between different membrane types in mitochondria. These analyses allow for the quantitative comparison of membrane features e.g. for different growth conditions. To demonstrate the utility of the toolkit tomogram datasets of mitochondria in the presence and absence of ER stress were analyzed. The authors conclude that ER stress affects mitochondria morphology through remodeling of the membrane structure.
The presented biological results and statistics are convincing and show active mitochondrial membrane remodeling in the cell when exposed to ER stress. It is also clear that there is a need for more quantitative evaluation based on the wealth of tomographic image features and mitochondrial membranes are certainly a well-chosen application. For this purpose, the authors developed a new workflow even though most of the discussed analyses are very specific to mitochondrial structures. Therefore, broader applications of these tools to other organelles are not easily envisaged without significant adaption. In that context, the title and abstract overpromise a much more powerful utility that can be applied to any other membrane analysis. Rather it seems that the proposed workflow is more of a specific tool or a pipeline for mitochondrial inner and outer membrane analysis instead of a toolkit for general morphological analysis. Hence, the manuscript cannot be accepted in its current form. In particular, the structure needs a significant rework of editing to become more comprehensible.

We appreciate the criticism that our workflow as implemented at the time of preprint is seemingly too focused on mitochondrial membranes and is not general. We've overhauled our workflow into a configurable (through a project YML file) scripted workflow that can take a folder with arbitrary segmentations and convert them into high quality meshes, followed by per-triangle quantification of the four primary metrics we describe in the manuscript: inter-membrane distance, intra-membrane through-space distance, curvature, and orientation. Generating fully automated visualization tools is more challenging, because which quantities are measured and how they are sub-classified (e.g., as we did for cristae, junctions, and IBM) is very project-specific; however, we did convert our visualization script into a library of utilities to combine tomograms into experiment objects, with methods to serialize for rapid access and functions for generating statistics and plots. Our converted visualizations script has been reorganized to act as an example of how similar questions could be asked for arbitrary membranes.

We propose to further demonstrate the generality of this updated approach by segmenting several examples of another organelle, the autophagosome, found in our dataset and applying the workflow to them in a supplementary figure.

4. The focussing to a method paper will also require more in-depth descriptions of the methodology in the main text. Although the code is deposited at github, there is no script-based workflow and description presented in the manuscript. Although Figure 1 puts the work into context of tomography, it remains very superficial on the image analysis. What are the input and output formats required for each step to follow the sequence of the workflow and at which steps critical interactive input is needed? What are the hardware requirements (CPU, GPU) or performance characteristics (CPU hours for certain operations)?

In addition to the changes mentioned above, we also added a "Supplemental Table 1" detailing computational requirements and time for each step.

We expanded on the description of this approach in the first paragraph of the results section:

"With this strategy, we were able to segment 32 tomograms containing mitochondria, divided between the elongated and fragmented bulk morphology populations and the two treatment groups (Figure 2, Supplementary Figure 1). The segmentation output was fed into the fully automated surface morphometrics pipeline (Figure 2B, Supplementary Figure 2, Supplementary Table 1). The voxel segmentation was converted to high quality membrane surfaces using the screened poisson algorithm[32]. Next, these surfaces were converted into triangle graphs and curvedness was estimated using pycurv[15], and the distances within and between surfaces as well as the relative orientations of different surfaces were estimated using the resulting graph. Finally, the quantifications for each tomogram were combined into experiments to allow aggregate statistics and visualizations. This 3D surface morphometrics pipeline is configurable for any segmented membrane and is available at https://github.com/grotjahnlab/surface_morphometrics."

We added a description of the up to date workflow in the methods section:

"Software workflow
The surface morphometrics pipeline is a python 3 scripted workflow with requirements that can be installed as a conda environment contained in an `environment.yml` file. The workflow is fully scripted and configurable with a `config.yml` file, and is run in 3 steps, with statistical analysis and visualization as an optional fourth step. First, a segmentation MRC file is converted automatically to a series of surface meshes formatted in the VTP file format. Second, for each mesh, the surface is converted to a graph (tg format) and curvature is estimated using pycurv. Third, orientations and distances between and within surfaces are calculated using the resulting graphs, and a CSV with quantifications as well as a final VTP surface file is output with all quantifications built in. Fourth, the outputs from multiple tomograms are combined for visualization and statistical analysis. Times and computational requirements are shown in supplementary table 1."

5. Figures 3-7 contain colorful 3D renderings of the measured quantities. In addition, they are filled with histograms of every possible quantitative parameter, which often are not very significant or different between. The authors should focus the main results and the figures to show the most relevant and significant findings and put the remaining panels and results into the supplement.
Figures 3-7 were organized around the different methodologies (inter and intra-membrane spacing, curvature, orientation) but we agree that focusing to the main results of each methodology is sufficient to show the value of these results. We propose to address this

criticism by moving figure 4D,F (inter-crista and junction spacing), figure 6 E,G (the junction measurements) and Figure 7 to supplemental figures. These supplemental figures will also be joined by the previously requested OMM curvature analysis and our proposed analysis of autophagosomes.

8. One of the key steps is the generation of a smooth surface from a segmented membrane, there is a question whether true membrane disruptions will be smoothed and may be overlooked in this approach. When these disruptions present true membrane ruptures, they may be of particular biological importance. The authors should support the choice and selection of the smoothing parameters in order to illustrate this potential pitfall.
The smoothing and hole-filling parameters are now configurable using the point_weight and extrapolation_voxels parameters in the config.yml file. Notably, the surfaces used for quantification used minimal smoothing, and any triangles more than a single voxel away from the point cloud were deleted, in order to ensure that the quantifications were minimally impacted by "hallucinated" surfaces. Additionally, the following text was added to the methods section discussion surface reconstruction:

"A surface mesh was calculated from the oriented point cloud using the screened Poisson algorithm[32], with a reconstruction depth of 9, an interpolation weight of 0.7, and a minimum number of samples of 1.5. These settings were chosen to maximize correspondence to the data, rather than smoothness. The resulting surface extended beyond the segmented region, so triangles more than 1 voxel away from the point cloud were deleted. Interpolation weight (point_weight) and the mask distance (extrapolation_voxel) are both configurable in the surface morphometrics pipeline if more aggressive smoothing and hole filling are desirable."

9. Throughout the manuscript, the authors mention statistical significance several times and one of the main aims of the study is perform statistical hypothesis testing. It is important to specify the significance test (not only in the methods) and the p-value in order to support this claim. In the manuscript, the authors use exclusively the Mann-Whitney test. What is the rationale for choosing this test? Have the authors considered comparing the total distributions and not just the peaks with e.g. a Kolmogorov-Smirnov test? For a statistical methods paper, there are also no discussion on error analysis.
This was a common concern raised by both reviewers, and we agree that a test based on total distribution would be more powerful than only looking at peaks. We address the use of the Kolmogorov-Smirnov test and the limitations we have run into thus far in our response to reviewer 1 in detail. In brief, KS tests tend to vastly overestimate statistical significance because the number of samples (the number of triangles) is vastly larger than the true number of independent features sampled in the data, so that even very similar looking distributions such as those in figure 5C yield p values in the range of 10^-200. We propose several approaches to better estimate the number of independent variables. We will also use a random subsampling approach within individual mitochondria to ensure sampling from the same distribution does not yield statistically significant results.

In addition to testing additional approaches to incorporate KS testing (based on estimation of number of independent features in each tomogram), we propose to improve our peak-based statistics by estimating a standard error for the peak of each tomogram using a bootstrap approach, getting the peaks from different random subsamples of triangles.

<u>Reviewer #2, minor comments:</u>

1. https://github.com/grotjahnlab/surface_morphometrics should include an example data set or tutorial for dissemination.
We are in the process of uploading all frame-averaged tilt series, tomograms, segmentations, and reconstructed surfaces to EMPIAR. Additionally, we propose to implement a complete tutorial including a single tomogram for readier workflow testing, separate from the complete data upload.

3. Description of the revisions that have already been incorporated in the transferred manuscript

Reviewer #1, major comments:

- As the work reported here is heavily computational, additional details about the computer hardware used and the time it took for the calculations to complete would be helpful for readers considering applying the code to their own data.
We appreciate the suggestion and included Supplementary Table 1 in the supplemental material outlining the computation time per step in our analysis pipeline:

"Supplemental Table 1. Approximate time and for each step of the surface morphometrics workflow.
Representative times and computational resources used for each step of the surface morphometrics workflow for each tomogram (unless otherwise noted) by the authors. Most time-intensive calculations were run in parallel on a compute cluster for each tomogram.

Step	Human Time (HH:MM)	Computational Wall Clock time (HH:MM)	CPU Cores Used	RAM Used
Automated initial segmentation (TomoSegMemTV)	00:10*	00:10*	8	64GB
Manual segmentation cleanup and classification	03:00	N/A	8	64GB
Point cloud conversion and mesh generation	00:01	00:03	4	16GB
Graph generation and curvature estimation (pycurv)	00:01	01:40	16	128GB
Distance and orientation measurement	00:01	00:10	16	128GB
Assembly of outputs from multiple tomograms into dataframes and serialization	00:01	00:10	1	16GB
Visualizations and statistical tests	00:01	00:10	1	16GB

* Tomosegmemtv is sometimes run iteratively with different settings to improve output. 10 minutes is approximately the time taken for a run without iteration, in the case of good output."

Reviewer #1, minor comments:

Revision Plan

- Pink and purple very close, consider alternative pair of colors or different shades to distinguish OMM and IMM
We kept OMM as purple but changed IMM to orange for Figure 3-7, and will make the associated changes to Figure 2 and Supplementary Movie 1 on final submission.

- Orientation of scaleboxes/scalebars should be consistent per figure panel. If knowledge of the axes is important to the reader, these should be included as well.
We followed the reviewer's suggestion and updated the scale cubes to be standardized per panel.

- In the last sentence of the introduction, the term "organellar architectures" is used, instead of the previously defined "membrane ultrastructure." Consider changing for clarity.
We changed "organellar architectures" to "membrane ultrastructure" in the last sentence of the abstract.

- Inconsistent use of the phrase "cryo-electron tomography" after defining and using "cryo-ET"
We changed all instances of "cryo-electron tomography" to "cryo-ET" after defining in the first instance in the introduction.

- Authors argue that the distinction between curvedness and curvature is important and that curvature is less appropriate in this context, but then use curvature in the abstract, throughout introduction and in the results section. Usage can be improved for readability.
We changed all instances of "curvature" to "curvedness" throughout the text and figure legends.

- In section "Development of a framework to automate quantification of ultrastructural features of cellular membranes" the second last sentence should read "... higher quality membrane surfaces as compared..."
We changed "surface" to "surfaces" in text.

- In section "IMM curvedness is differentially sensitive to Tg treatment in elongated and fragmented mitochondrial networks" the fourth sentence should perhaps read "... despite apparent visual differences, no significant..."
We changed "difference" to "differences" in text.

- The term "cell's growth plane" is not clear from the text nor from Fig 6A. Do the authors mean surface of the substrate the cell is growing on?
We clarified and further defined the "cell's growth plane" in the text by adding the following phrase:

"… the cell's growth plane (i.e. the plane of electron microscopy grid substrate to which the cell is adhered) (Figure 6A)."

- In Materials and Methods:
- The authors report that manual back-blotting was used in a Vitrobot. This is non-standard usage and more details should be provided.
We added the following description to clarify our manual back-blotting procedure on the Vitrobot:

"After 8 hours of incubation, samples were plunge-frozen in a liquid ethane/propane mixture using a Vitrobot Mark 4 (Thermo Fisher Scientific). The Vitrobot was set to 37° C and 100% relative humidity and blotting was performed manually from the back side of grids using Whatman #1 filter paper strips through the Vitrobot humidity/temperature chamber side port. The Vitrobot settings used to disable automated blotting apparatus were as follows: Blot total: 0, 2; Blot force: 0, 3; Blot time: 0 seconds."

- In section "Fluorescence Guided Milling" in the third sentence, the word "based" is repeated, second can be removed.
We deleted the second instance of "based" in this sentence.

- Symbol for degree (or the word degree) should be added to angular increment and tilt range for clarity.
Added degree symbols to the following sentence in the "Tilt Series Data Collection" portion of the materials and methods:

"Tilt series were acquired using SerialEM software (Mastronarde, 2005) with 2° steps between -60° and +60°."

- Capitalization of TomoSegMemTV is inconsistent.
We changed all mentions to TomoSegMemTV.

- Fig 3 title - consider replacing "Inter-mitochondrial membrane..." with "Intra-mitochondrial membrane..." for clarity.
We clarified this point by changing "Inter-mitochondrial membrane distance" to "Distance between inner and outer mitochondrial membranes" in the figure legend:

"Figure 3. Distance between inner and outer mitochondrial membranes is dependent on mitochondrial network morphology and presence or absence of ER stress."

- Fig 3C caption - should explicitly state it is a combined histogram and that the dashed lines correspond to the peak of the pooled data.
We changed "Quantification of" to "Combined histogram of" and added the sentence " to each of the relevant figure captions (Fig. 3c, 4c-f, 5b-e, 6d-g, 7c):

"Dashed vertical lines correspond to peak histogram values of pooled data"

- Fig 6B and 6C caption - upper and lower parts not explicitly described.
We modified Fig 6B&C caption to more clearly describe the figure panel:

"(B) Two representative membrane surface reconstructions of lamellar Tg-treated elongated mitochondria, colored by angle of IMM relative to OMM.

(C) Two representative membrane surface reconstructions of a less rigidly oriented Tg-treated elongated mitochondria, colored by angle of IMM relative to the growth plane of the cell."

<u>Reviewer #2, major comments:</u>

1. Title and abstract need to be toned down not to overpromise a very general toolkit. The presented method may be a tool or a collection of scripts - a toolkit can be used to address

other types of (membrane) analysis problems. In the end, the analysis builds to a large extent on the previous developments and implementation of PyCurve. Perhaps, the most interesting contribution here is the application of the mesh generation by the Poisson reconstruction method to the segmented membranes, which is, however, well implemented in the used pymeshlab framework. The computation of distances and angles is straightforward.
We appreciate this critique and do not want to overpromise with our work, although we believe the overhaul to a fully configurable workflow addresses the primary concern. We are quite clear in the text that we build on top of pycurv, and recommend citation of the original tool as well as our pipeline in the github repository as a result. With that said,

We have changed the title as follows:

"Quantifying mitochondrial ultrastructure in cryo-electron tomography using a surface morphometrics pipeline"

We have also renamed our method to the surface morphometrics pipeline to reduce over-implication of generality, and made other small changes to increase degree of detail about what our method is resolving.

2. When reading the manuscript, the reader is left in the open whether this is a method paper or a biological results paper. The title/abstract suggests that this is a method paper and the manuscript is more of a mitochondrial membrane report in ER stress. Therefore, the title/abstract does not reflect the manuscript very well.
We aim to use this manuscript to describe the development of a workflow that enabled novel and interesting biological results. We adjusted the title to better match the combined development of a new pipeline and application to an interesting biological system as proof of concept:

"Quantifying mitochondrial ultrastructure in cryo-electron tomography using a surface morphometrics pipeline"

3. The manuscript also requires substantial structural editing. Several references to Figures are not appearing in the text in the order that the Figure panels are built. Excessive cross-referencing of figures also make the manuscript hard to read.
We simplified our referencing of figures and made sure the text matched the order of the figure panels.

6. The exact morphological discrimination between fragmented and elongated mitochondria is not easily understood from the results section. What is really meant by blinded manual classification? It only became clear when reading the methods. The results section should stand on its own. How is the overall population between fragmented and elongated cells is affected after Tg application?
To clarify our methodology for blinded classification of mitochondrial network morphologies we included the following text:

"We categorized cells for mitochondrial network morphology by blinded manual classification in which five researchers were given fluorescence microscopy images of exemplar network morphologies (elongated and fragmented) as references to assign morphologies to the experimental fluorescence micrographs."

We targeted similar ratios of elongated and fragmented cells in both vehicle and Tg treated conditions for tomography, but qualitatively saw the expected increase in the elongated population to what has been previously described during Tg treatment. Because of our single cell targeting approach we did not quantify the population shift."

7. Similarly, what is meant by manual classification of IMM, OMM and ER? Is there any clustering involved?
Our automated segmentation approach labels all membranes, and the separation of the IMM, OMM, and ER membranes is done by an expert user selecting and relabeling each membrane based on cellular context (e.g. IMM is inside of OMM and contains cristae). We have added the following text to clarify our methodology for manual classification of IMM, OMM, and ER:

"This was followed by manual labeling of membranes into mitochondrial IMM and OMM and ER membrane based on cellular context, as well as manual cleanup of individual membrane segmentations using AMIRA software (Thermo Fisher Scientific)."

Reviewer #2, minor comments:
2. What is meant by growth plane? This term is not defined in the manuscript.
We clarified and further defined the "cell's growth plane" in the text by adding the following phrase:

"… the cell's growth plane (the plane of electron microscopy grid substrate on which the cell is grown) (Figure 6A)."

3. What is meant by vehicle treatment? There is no explanation in the main text of the manuscript.
We clarified and further defined vehicle treatment in the main text by adding the following:

"We applied our correlative approach to identify and target specific Tg-treated and vehicle (media with DMSO) treated MEF^{mtGFP} cells with either elongated or fragmented mitochondrial network morphologies for cryo-FIB milling and cryo-ET data acquisition and reconstruction."

5. Have the authors noticed/calculated any differences in the width of the cristae?
We measure this difference in figure 4C (Intra-crista distance). We found significant changes in width/intra-crista distance in response to Tg treatment in both elongated and fragmented morphologies.

6. Methods: Automated surface reconstruction: "In cases where the resulting surface was very complex, the surface was simplified..." How was the complexity determined?
With the updated state of the software, we simplify all surfaces to generate a maximum of 150,000 triangles. This has minimal effect on very small surfaces, but greatly speeds computation on very large surfaces. We corrected the language to match this:

"The resulting mesh was simplified with quadric edge collapse decimation to produce a surface that represented the membrane with 150,000 triangles or fewer."

7. Methods: Calculation of distances between individual surfaces: "For surfaces with small numbers of triangles, this was accomplished using a distance matrix...". What is the threshold for a small number of triangles?

As part of our software overhaul we have changed to always using a more memory-efficient KD tree based quantification, since the additional speed for the distance matrix approach is minimal when there are few enough triangles for it to be appropriate, and the hardwired cutoff was not as flexible for different hardware configurations. The updated text is below, but to satisfy any potential reviewer curiosity, the decision was made when the required distance matrix would use more than 128GB of memory. In the case of two identically sized surfaces, this crossover happens when there are approximately 45,000 triangles in each surface.

"For calculations of distances between respective surface meshes, the minimum distance from each triangle on one surface to the nearest triangle on the other surface was calculated using a KD-tree."

Reviewer #2 (Significance (Required)):

The aim of the paper is well motivated. Cryo-ET is a growth field and there is a need for quantitative parameterization of cryo-ET data. Recently a toolkit for the analysis of filaments from cryo-ET has been published (Dimchev et al. 2021 DOI: 10.1016/j.jsb.2021.107808). Given the specific nature of the implementation, i.e. the membrane structures of mitochondria, I cannot easily see that this implementation will be useful beyond the analysis of mitochondrial membrane structure.
We hope that we have addressed this concern with generality has been addressed by our previously described updates to the software implementation.

4. **Description of analyses that authors prefer not to carry out.**

Review 2, minor comments:
4. Angle between OMM and cristae: Maybe use the average angle of each cristae for comparison or fit a plane for each cristae because you are interested in the angle between the cristae and the OMM and the membrane of the cristae has a lot of uneven surfaces
We believe that the advantage of our approach is the ability to incorporate more complex geometric information from uneven surfaces such as those seen in cristae. With that said, the ability to quantify metrics for individual cristae in an automated manner would be very appealing, since in many ways cristae are functionally independent compartments. Accomplishing this would require either subdividing the larger surface into individual cristae, which will require development of additional sub-graph processing strategies. Additionally, pairing surfaces to represent opposite sides of a crista will require additional development. While we agree that this will be an excellent extension of the surface morphometrics approach, we feel that the additional development required is out of the scope of this initial manuscript focused on the general workflow. New methods leveraging sub-graph analysis will be explored in future manuscripts.

May 8, 2022

Re: JCB manuscript #202204093T

Dr. Danielle Ann Grotjahn
Scripps Research Institute
Department of Integrative Structural and Computational Biology
10550 North Torrey Pines Rd
Hazen-173
La Jolla, California 92037

Dear Dr. Grotjahn,

Thank you for submitting your manuscript "Quantifying mitochondrial ultrastructure in cryo-electron tomography using a surface morphometrics pipeline." We have now had an opportunity to assess your manuscript, the reviewer reports from Review Commons, and your revision plan.

We agree with the reviewers that your new workflow could represent an important and significant advance. However, we do not believe that the study would be a good fit for JCB without a significant extension that demonstrates general utility of the methodology beyond just analyzing mitochondrial network morphology. Your plan to analyze autophagosomes would be sufficient to address this concern but this would also require a substantial amount of additional data and analyses on your part. If you are able to add this, then we would be open to re-reviewing a revised version of this work. Please note that in addition to the original Review Commons reviewers we will likely recruit a third reviewer to assess the entire manuscript and the statistical analyses. We also feel that additional discussion is necessary to properly place the observed changes in mitochondrial morphology under ER stress conditions in context of prior studies and to better explain the biological significance of these findings.

If you are not able to undertake these significant revisions and would like to submit your paper to another Review Commons affiliate journal, please let us know and we can inform the Review Commons office so they can release the paper in their system.

Please let us know if you are able to address the major issues outlined above and wish to submit a revised manuscript to JCB. Note that a substantial amount of additional experimental data likely would be needed. The typical timeframe for revisions is three to four months. While most universities and institutes have reopened labs and allowed researchers to begin working at nearly pre-pandemic levels, we at JCB realize that the lingering effects of the COVID-19 pandemic may still be impacting some aspects of your work, including the acquisition of equipment and reagents. Therefore, if you anticipate any difficulties in meeting this aforementioned revision time limit, please contact us and we can work with you to find an appropriate time frame for resubmission. Please note that papers are generally considered through only one revision cycle, so any revised manuscript will likely be either accepted or rejected.

If you choose to revise and resubmit your manuscript, please also attend to the following editorial points. Please direct any editorial questions to the journal office.

GENERAL GUIDELINES:
Text limits: Character count is < 40,000, not including spaces. Count includes title page, abstract, introduction, results, discussion, and acknowledgments. Count does not include materials and methods, figure legends, references, tables, or supplemental legends.

Figures: Your manuscript may have up to 10 main text figures. To avoid delays in production, figures must be prepared according to the policies outlined in our Instructions to Authors, under Data Presentation, https://jcb.rupress.org/site/misc/ifora.xhtml. All figures in accepted manuscripts will be screened prior to publication.

Materials and methods: Should be comprehensive and not simply reference a previous publication for details on how an experiment was performed. Please provide full descriptions (at least in brief) in the text for readers who may not have access to referenced manuscripts. The text should not refer to methods "...as previously described." For all cell lines, vectors, constructs - all genetic material: please include database / vendor ID (e.g., Addgene, ATCC, etc.) or if unavailable, please briefly describe their basic genetic features, even if described in other published work or gifted to you by other investigators (and provide references where appropriate). Please add details regarding the MEF cell line and the mitochondria-targeted GFP.

Statistical analysis: Error bars on graphic representations of numerical data must be clearly described in the figure legend. The number of independent data points (n) represented in a graph must be indicated in the legend. Statistical methods should be

explained in full in the materials and methods. For figures presenting pooled data the statistical measure should be defined in the figure legends. Please also be sure to indicate the statistical tests used in each of your experiments (both in the figure legend itself and in a separate methods section) as well as the parameters of the test (for example, if you ran a t-test, please indicate if it was one- or two-sided, etc.). Also, if you used parametric tests, please indicate if the data distribution was tested for normality (and if so, how). If not, you must state something to the effect that "Data distribution was assumed to be normal but this was not formally tested."

IMPORTANT: It is JCB policy that if requested, original data images must be made available. Failure to provide original images upon request will result in unavoidable delays in publication. Please ensure that you have access to all original microscopy and blot data images before submitting your revision.

Supplemental information: There are strict limits on the allowable amount of supplemental data. Your manuscript may have up to 5 supplemental figures. Up to 10 supplemental videos or flash animations are allowed. A summary of all supplemental material should appear at the end of the Materials and methods section.

Please note that JCB now requires authors to submit Source Data used to generate figures containing gels and Western blots with all revised manuscripts. This Source Data consists of fully uncropped and unprocessed images for each gel/blot displayed in the main and supplemental figures. If your paper will include cropped gel and/or blot images, please be sure to provide one Source Data file for each figure that contains gels and/or blots along with your revised manuscript files. File names for Source Data figures should be alphanumeric without any spaces or special characters (i.e., SourceDataF#, where F# refers to the associated main figure number or SourceDataFS# for those associated with Supplementary figures). The lanes of the gels/blots should be labeled as they are in the associated figure, the place where cropping was applied should be marked (with a box), and molecular weight/size standards should be labeled wherever possible. Source Data files will be made available to reviewers during evaluation of revised manuscripts and, if your paper is eventually published in JCB, the files will be directly linked to specific figures in the published article.

Source Data Figures should be provided as individual PDF files (one file per figure). Authors should endeavor to retain a minimum resolution of 300 dpi or pixels per inch. Please review our instructions for export from Photoshop, Illustrator, and PowerPoint here: https://rupress.org/jcb/pages/submission-guidelines#revised

If you choose to resubmit, please include a cover letter addressing the reviewers' comments point by point. Please also highlight all changes in the text of the manuscript.

Regardless of how you choose to proceed, we hope that the comments below will prove constructive as your work progresses. We would be happy to discuss them further once you've had a chance to consider the points raised. You can contact the journal office with any questions, cellbio@rockefeller.edu or call (212) 327-8588.

Thank you for thinking of JCB as an appropriate place to publish your work.

Sincerely,

Richard Youle, PhD
Monitoring Editor
Journal of Cell Biology

Dan Simon, PhD
Scientific Editor
Journal of Cell Biology

Revision Plan

Manuscript number: RC-2022-01274
Corresponding author(s): Danielle Grotjahn

Blue text: revision plan discussed with Journal of Cell Biology editors
Red text: addressed revisions for revised submission to Journal of Cell Biology

1. General Statements [optional]

We would like to thank the reviewers for their prompt and thoughtful input on our manuscript, and their willingness to participate in more portable review through Review Commons.

2. Description of the planned revisions

We have addressed and incorporated all suggested revisions in the transferred manuscript.

3. Description of the revisions that have already been incorporated in the transferred manuscript

Reviewer #1, major comments:
- A major concern is that the data are reported and analyzed on a per tomogram basis when many tomograms contain multiple mitochondria. Given that the mitochondria appear mostly well separated in Sup. Fig 1 with only a few connections visible, and the high degree of pleomorphism noted by the authors, I would strongly suggest that the authors use each mitochondrion as the basis for reporting their metrics rather than the FOV/tomogram as this would avoid mixing metrics from different mitochondria that may be in different states (e.g., fusion/fission). This would apply to data shown in Figures 3, 4, 5, and 6.

Proposed Revision Plan: We appreciate the reviewer suggestion to separate on a per mitochondrion vs per tomogram basis for our analysis. While we do not anticipate that this will significantly change the overall findings, we agree that splitting per mitochondrion will account for any possible variability between mitochondria within the given field of view. Furthermore, we anticipate that this will improve our analysis and statistical power by effectively increasing the total sample size per experimental group. For our next revision, we will divide surfaces on a per mitochondrion basis within a given tomogram and re-run the full analysis pipeline. Additionally, per reviewer request, we will include an output histogram for each measurement per mitochondrion surface in a supplemental figure.

*How this was addressed in the revised manuscript: We divided surfaces and performed our analyses on a per mitochondrion basis. As expected, this improved our analysis and statistical power by effectively increasing the total sample size per group, with limited effect on the overall results. We output histograms for each measurement per mitochondrion surface. We included a representative example of the per mitochondrion histograms for the OMM-IMM distances in Supplemental Figure 3. Given the large quantity of analyses produced throughout this manuscript, we opted to deposit the full dataset of per mitochondrion analyses for each measurement in Zenodo (*10.5281/zenodo.7102596*). We noted this in page 38, lines 906-907:*

"*The complete output of the per mitochondrion analyses for every measurement is available at*

10.5281/zenodo.7102596."

Revision Plan

- In Figure 3C the authors show the combined distribution of OMM-IMM distances within each condition. This may obscure some variability within populations. Individual histograms for all mitochondria should be included as supplementary material. Currently, it is difficult to judge if the peak of the combined distribution is appropriate and impossible to judge the variability between tomograms (preferably mitochondria, see above comment). Additionally, the shape of the distributions appears significantly different between conditions, suggesting that selecting a single peak value as representative and the basis for the statistical tests (Fig 3D) might not be appropriate. Please comment.

Proposed Revision Plan: We will include individual histograms for each measurement per mitochondrion surface in a supplemental figure.

We agree that peak-based statistical tests limit our ability to quantify more complex differences, and therefore we chose to output histograms in addition to violin plots, so that shape differences can be observed qualitatively. A major challenge of shape-based statistical quantification is the assessment of independent samples. By using peak-based quantification, we could assume that each tomogram (and in the planned revision, each mitochondrion) is an independent sample, but for shape distribution this is inappropriate since there is more than one value represented per tomogram. Running a KS test with N equal to the number of tomograms yields no significance even in the visible cases where the shape appears very different.

However, the number of triangles also poorly represents the number of independent samples, since 1) the number of triangles used to represent a surface is somewhat arbitrary and remeshing can change it dramatically and 2) Our chosen triangle size is considerably smaller than the visually observed feature size in order to allow effective vector voting in the pycurv AVV algorithm. The result of this is that when we use a KS test on the distribution of values per triangle, even visually identical distributions yield p-values below 10^-200.

We do estimate the approximate smallest feature size during our calculations, since that is used to generate the radius used by pycurv in vector voting, to be 12 nm (the radius hit parameter in pycurv). During a public presentation of this work an audience member suggested that we might use the area implied by this feature size (~450 nm^2) as the size of an independent sample. This would yield around 1000 independent samples per tomogram. Because the choice of feature size is heuristic and manual, this is not as statistically sound as the peak-based metric, which is why we believe that the more conservative peak-based statistical testing is the gold standard for proving differences, but we believe this will be the most reliable way to quantify differences in shape of distributions. We plan to implement this quantification in our revision, and will evaluate whether it gives “expected” statistical results by a bootstrapping approach using subsampling of triangles from the same vs different mitochondria.

We would welcome reviewer suggestions for additional shape-based metrics and will explore other potential metrics to capture shape as part of our revision. While our peak-based metrics demonstrate our ability to statistically capture small changes in ultrastructure with this method, shape-based quantification will significantly enhance the capability to capture finer changes in structure that may be critical to understand physiologically.

Once this additional testing is complete, we will add a section to the results section describing choice of statistical framework. We also plan to generate a supplementary table showing the results of the peak-based quantification alongside all shape-based quantifications.

Revision Plan

How this was addressed in the revised manuscript: We have expanded the statistical frameworks used to include KS testing comparison of histogram peak sets to identify changes in the variance between mitochondria, Mann Whitney U testing of standard deviations rather than histogram peaks for orientation measurements, as well as KS testing of overall distributions to measure changes to the shapes of the distribution as a whole. We tested 3 methods of estimating the number of independent samples for this distribution KS testing (number of triangles, number of mitochondria, or number of "features" as estimated by dividing total membrane area by a feature area of 452 nm^2 corresponding to the 12nm feature radius used for vector voting. For the most part we still relied on the Mann Whitney U test for interpreting results in the examples used in the text but have added the additional statistical tests to the pipeline for cases where they can more clearly be distinguished. The details of these measurements are included in Supplementary Table 2, and the following text was added to the main text on pages 8-9, lines 119-229 in the section titled "*"Bulk analysis and assessment of statistical significance based on surface quantifications".*

- In Figure 4C-F, again combined distributions are shown. Authors should include individual histograms for all mitochondria as supplementary material. The diversity of distributions in the metrics are more pronounced than the distances in reported in Fig 3, again making assessment of variability difficult and raising doubt about using the single peak value.

Proposed Revision Plan: We will include individual histograms for each measurement per mitochondrion surface in a supplemental figure.

*How this was addressed in the revised manuscript: We included a representative example of the per mitochondrion histograms for the OMM-IMM distances in Supplemental Figure 3. Given the large quantity of analyses produced throughout this manuscript, we opted to deposit the full dataset of per mitochondrion analyses for each measurement in Zenodo (*10.5281/zenodo.7102596*). We noted this in page 38, lines 906-907:*

"*The complete output of the per mitochondrion analyses for every measurement is available at 10.5281/zenodo.7102596*."

As we describe above, we will test several options for distribution-based statistical quantifications and incorporate the results in the manuscript. We expect them to be useful for every measurement we make.

We incorporated the additional statistical frameworks described above and the measurements for this figure and others are available in Supplemental Table 2.

- It would be helpful to include the curvature or curvedness of the OMM for each mitochondrion in the supplementary material. The data to correlate OMM curvature with elongated/fragmented mitochondria should be available and might be of interest to some readers.

Proposed Revision Plan: We will calculate curvedness of the OMM for each mitochondrion and include these data in the supplemental material. The inverse of the curvedness of the OMM gives a reasonable approximation of the radius of the mitochondrial "tube", a feature which can be challenging to quantify fully automatically, and we agree that this may be of particular interest to some of our readers – particularly if morphology changes or stress-driven changes alter that radius in a statistically significant way.

Revision Plan

How this was addressed in the revised manuscript: We calculated curvedness of the OMM for each mitochondrion and included these data in Supplemental Figure 4. We noted this in the text in page 12 and lines 304-308.

"OMM curvedness is dependent on mitochondrial network morphology
We calculated curvedness of the OMM for each mitochondrion. For both vehicle and Tg treated cells, curvedness of the OMM is decreased in fragmented mitochondria relative to elongated networks (Supplemental Figure 4). This change reflects the visual observation that fragmented mitochondria appear wider and more spherical relative to elongated mitochondria which appear more tubular in shape with higher curvature regions."

- As the work reported here is heavily computational, additional details about the computer hardware used and the time it took for the calculations to complete would be helpful for readers considering applying the code to their own data.

Proposed Revision Plan: We appreciate the suggestion and included Supplementary Table 1 in the supplemental material outlining the computation time per step in our analysis pipeline.

How this was addressed in the revised manuscript: Please view Supplementary Table 1 in the revised manuscript.

Reviewer #1, minor comments:

- For all data, exact n per condition should be given (in text and captions as appropriate), not a range for the whole set.

Proposed Revision Plan: We will report the exact n per condition in text and in captions after we separate our data on a per mitochondrion basis and update the analysis.

How this was addressed in the revised manuscript: We reported the exact n (number of mitochondria per condition) in the figure legends for Figures 3-6 and Supplemental Figures 4, 5, 8.

- Fig 5E middle, legend obscures some of the data.

Proposed Revision Plan: We will reformat the graph such that the legend does not obscure the data after we separate our data on a per mitochondrion basis and update the analysis.

How this was addressed in the revised manuscript: We reformatted all the new histograms after separating our data on a per mitochondrion basis and verified that the figure legends do not obscure data.

- Pink and purple very close, consider alternative pair of colors or different shades to distinguish OMM and IMM

Proposed Revision Plan: We kept OMM as purple but changed IMM to orange for Figure 3-7, and will make the associated changes to Figure 2 and Supplementary Movie 1 on final submission.

Revision Plan

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- Orientation of scaleboxes/scalebars should be consistent per figure panel. If knowledge of the axes is important to the reader, these should be included as well.

Proposed Revision Plan: We followed the reviewer's suggestion and updated the scale cubes to be standardized per panel.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- In the last sentence of the introduction, the term "organellar architectures" is used, instead of the previously defined "membrane ultrastructure." Consider changing for clarity.

Proposed Revision Plan: We changed "organellar architectures" to "membrane ultrastructure" in the last sentence of the abstract.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- Inconsistent use of the phrase "cryo-electron tomography" after defining and using "cryo-ET"

Proposed Revision Plan: We changed all instances of "cryo-electron tomography" to "cryo-ET" after defining in the first instance in the introduction.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- Authors argue that the distinction between curvedness and curvature is important and that curvature is less appropriate in this context, but then use curvature in the abstract, throughout introduction and in the results section. Usage can be improved for readability.

Proposed Revision Plan: We changed all instances of "curvature" to "curvedness" throughout the text and figure legends.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- In section "Development of a framework to automate quantification of ultrastructural features of cellular membranes" the second last sentence should read "… higher quality membrane surfaces as compared".."

Proposed Revision Plan: We changed "surface" to "surfaces" in text.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

- In section "IMM curvedness is differentially sensitive to Tg treatment in elongated and fragmented mitochondrial networks" the fourth sentence should perhaps read "… despite apparent visual differences, no significant.."

Proposed Revision Plan: We changed "difference" to "differences" in text.

How this was addressed in the revised manuscript: This was completed in the revised manuscript.

Revision Plan

- The term “cell’s growth plane” is not clear from the text nor from Fig 6A. Do the authors mean surface of the substrate the cell is growing on?
Proposed Revision Plan: We clarified and further defined the “cell’s growth plane” in the text by adding the following phrase:

“… the cell’s growth plane (i.e. the plane of electron microscopy grid substrate to which the cell is adhered) (Figure 6A).”

How this was addressed in the revised manuscript: This is included in the revised manuscript.

- In Materials and Methods:
- The authors report that manual back-blotting was used in a Vitrobot. This is non-standard usage and more details should be provided.

Proposed Revision Plan: We added the following description to clarify our manual back-blotting procedure on the Vitrobot:

“*After 8 hours of incubation, samples were plunge-frozen in a liquid ethane/propane mixture using a Vitrobot Mark 4 (Thermo Fisher Scientific). The Vitrobot was set to 37° C and 100% relative humidity and blotting was performed manually from the back side of grids using Whatman #1 filter paper strips through the Vitrobot humidity/temperature chamber side port. The Vitrobot settings used to disable automated blotting apparatus were as follows: Blot total: 0, 2; Blot force: 0, 3; Blot time: 0 seconds*.”

How this was addressed in the revised manuscript: This is included in the revised manuscript.

- In section "Fluorescence Guided Milling" in the third sentence, the word "based" is repeated, second can be removed.
Proposed Revision Plan: We deleted the second instance of “based” in this sentence.

How this was addressed in the revised manuscript: This change is made in the revised manuscript.

- Symbol for degree (or the word degree) should be added to angular increment and tilt range for clarity.

Proposed Revision Plan: Added degree symbols to the following sentence in the “Tilt Series Data Collection” portion of the materials and methods:

“*Tilt series were acquired using SerialEM software (Mastronarde, 2005) with 2° steps between -60° and +60°.”*

How this was addressed in the revised manuscript: This change is made in the revised manuscript.

- Capitalization of TomoSegMemTV is inconsistent.

Proposed Revision Plan: We changed all mentions to TomoSegMemTV.

How this was addressed in the revised manuscript: This change is made in the revised manuscript.

- Fig 3 tit–e - consider replacing "Inter-mitochondrial membrane”.." with "Intra-mitochondrial membrane”.." for clarity.

Proposed Revision Plan: We clarified this point by changing “Inter-mitochondrial membrane distance” to “Distance between inner and outer mitochondrial membranes” in the figure legend:

“*Figure 3. Distance between inner and outer mitochondrial membranes is dependent on mitochondrial network morphology and presence or absence of ER stress.”*

How this was addressed in the revised manuscript: This change is made in the revised manuscript.

- Fig 3C caption - should explicitly state it is a combined histogram and that the dashed lines correspond to the peak of the pooled data.

Proposed Revision Plan: We changed “Quantification of” to “Combined histogram of” and added the sentence ” to each of the relevant figure captions (Fig. 3c, 4c-f, 5b-e, 6d-g, 7c):

“*Dashed vertical lines correspond to peak histogram values of pooled data*”

How this was addressed in the revised manuscript: This change is made in the revised manuscript.

- Fig 6B and 6C caption - upper and lower parts not explicitly described.
Proposed Revision Plan: We modified Fig 6B&C caption to more clearly describe the figure panel:

“*(B) Two representative membrane surface reconstructions of lamellar Tg-treated elongated mitochondria, colored by angle of IMM relative to OMM.*

(C) Two representative membrane surface reconstructions of a less rigidly oriented Tg-treated elongated mitochondria, colored by angle of IMM relative to the growth plane of the cell.”

How this was addressed in the revised manuscript: These changes are made in the revised manuscript.

Reviewer #2, major comments:

Barad, Medina et al. presents a new toolkit for the analysis of membrane ultrastructure in cryo-tomograms. More specifically, the toolkit is designed to compare curvature, angles and spacing between different membrane types in mitochondria. These analyses allow for the quantitative comparison of membrane features e.g. for different growth conditions. To demonstrate the utility of the toolkit tomogram datasets of mitochondria in the presence and absence of ER stress were analyzed. The authors conclude that ER stress affects mitochondria morphology through remodeling of the membrane structure. The presented biological results and statistics are convincing and show active mitochondrial membrane remodeling in the cell when exposed to ER stress. It is also clear that there is a need for more quantitative evaluation based on the wealth of tomographic image features and mitochondrial membranes are certainly a well-chosen

application. For this purpose, the authors developed a new workflow even though most of the discussed analyses are very specific to mitochondrial structures. Therefore, broader applications of these tools to other organelles are not easily envisaged without significant adaption. In that context, the title and abstract overpromise a much more powerful utility that can be applied to any other membrane analysis. Rather it seems that the proposed workflow is more of a specific tool or a pipeline for mitochondrial inner and outer membrane analysis instead of a toolkit for general morphological analysis. Hence, the manuscript cannot be accepted in its current form. In particular, the structure needs a significant rework of editing to become more comprehensible.

Proposed Revision Plan: We appreciate the criticism that our workflow as implemented at the time of preprint is seemingly too focused on mitochondrial membranes and is not general. We've overhauled our workflow into a configurable (through a project YML file) scripted workflow that can take a folder with arbitrary segmentations and convert them into high quality meshes, followed by per-triangle quantification of the four primary metrics we describe in the manuscript: inter-membrane distance, intra-membrane through-space distance, curvature, and orientation. Generating fully automated visualization tools is more challenging, because which quantities are measured and how they are sub-classified (e.g., as we did for cristae, junctions, and IBM) is very project-specific; however, we did convert our visualization script into a library of utilities to combine tomograms into experiment objects, with methods to serialize for rapid access and functions for generating statistics and plots. Our converted visualizations script has been reorganized to act as an example of how similar questions could be asked for arbitrary membranes.

We propose to further demonstrate the generality of this updated approach by segmenting several examples of another organelle, the autophagosome, found in our dataset and applying the workflow to them in a supplementary figure.

How this was addressed in the revised manuscript: We segmented and generated surfaces of three examples of autophagosome membranes present in our dataset and applied the workflow to these membrane surfaces. We included these results in Supplemental Figures 6 and 7 and noted this in the text in page 13 lines 333-343:

''Application of surface morphometrics and analysis to additional organellar membranes
We further demonstrated the generality of this approach by generating mesh surfaces for membranes of two other organelles present in our data: the endoplasmic reticulum (ER) (Supplemental Figure 1) and the autophagosome (Supplemental Figure 6). We applied our morphometrics pipeline to quantify parameters of membrane architecture of these organelles. Like mitochondria, autophagosomes are comprised of multiple membranes, including the outer autophagosomal membrane (OAM) and the inner autophagosomal membrane (IAM), as well as membranes from the engulfed cargo (cargo membranes) (Supplemental Figure 7). We then visualized the spatial distributions of the curvatures, distances, and orientations of these distinct membranes on the generated surface reconstructions (Supplemental Figure 7). Additionally, we calculated the curvedness of ER membranes in cells with differing mitochondrial populations and observed no significant changes in any of the conditions (Supplemental Figure 8)."

1. Title and abstract need to be toned down not to overpromise a very general toolkit. The presented method may be a tool or a collection of scripts - a toolkit can be used to address other types of (membrane) analysis problems. In the end, the analysis builds to a large extent on the previous developments and implementation of PyCurve. Perhaps, the most interesting contribution here is the application of the mesh generation by the Poisson reconstruction method

to the segmented membranes, which is, however, well implemented in the used pymeshlab framework. The computation of distances and angles is straightforward.

Proposed Revision Plan: We appreciate this critique and do not want to overpromise with our work, although we believe the overhaul to a fully configurable workflow addresses the primary concern. We are quite clear in the text that we build on top of pycurv, and recommend citation of the original tool as well as our pipeline in the github repository as a result. With that said,

We have changed the title as follows:

"*Quantifying organellar ultrastructure in cryo-electron tomography using a surface morphometrics pipeline*"

We have also renamed our method to the surface morphometrics pipeline to reduce over-implication of generality, and made other small changes to increase degree of detail about what our method is resolving.

How this was addressed in the revised manuscript: These changes are included in the revised submission.

2. When reading the manuscript, the reader is left in the open whether this is a method paper or a biological results paper. The title/abstract suggests that this is a method paper and the manuscript is more of a mitochondrial membrane report in ER stress. Therefore, the title/abstract does not reflect the manuscript very well.

Proposed Revision Plan: We aim to use this manuscript to describe the development of a workflow that enabled novel and interesting biological results. We adjusted the title to better match the combined development of a new pipeline and application to an interesting biological system as proof of concept:

"*Quantifying organellar ultrastructure in cryo-electron tomography using a surface morphometrics pipeline*"

How this was addressed in the revised manuscript: This change is included in the revised submission.

3. The manuscript also requires substantial structural editing. Several references to Figures are not appearing in the text in the order that the Figure panels are built. Excessive cross-referencing of figures also make the manuscript hard to read.

Proposed Revision Plan: We simplified our referencing of figures and made sure the text matched the order of the figure panels.

How this was addressed in the revised manuscript: These changes are included in the revised submission.

4. The focusing to a method paper will also require more in-depth descriptions of the methodology in the main text. Although the code is deposited at github, there is no script-based workflow and description presented in the manuscript. Although Figure 1 puts the work into context of tomography, it remains very superficial on the image analysis. What are the input and output

formats required for each step to follow the sequence of the workflow and at which steps critical interactive input is needed? What are the hardware requirements (CPU, GPU) or performance characteristics (CPU hours for certain operations)?

Proposed Revision Plan: In addition to the changes mentioned above, we also added a "Supplemental Table 1" detailing computational requirements and time for each step.

We expanded on the description of this approach in the first paragraph of the results section:

"*With this strategy, we were able to segment 32 tomograms containing mitochondria, divided between the elongated and fragmented bulk morphology populations and the two treatment groups (Figure 2, Supplementary Figure 1). The segmentation output was fed into the fully automated surface morphometrics pipeline (Figure 2B, Supplementary Figure 2, Supplementary Table 1). The voxel segmentation was converted to high quality membrane surfaces using the screened poisson algorithm[32]. Next, these surfaces were converted into triangle graphs and curvedness was estimated using pycurv[15], and the distances within and between surfaces as well as the relative orientations of different surfaces were estimated using the resulting graph. Finally, the quantifications for each tomogram were combined into experiments to allow aggregate statistics and visualizations. This 3D surface morphometrics pipeline is configurable for any segmented membrane and is available at https://github.com/grotjahnlab/surface_morphometrics*."

We also added a description of the up-to-date workflow in the methods section:

"Software workflow
The surface morphometrics pipeline is a python 3 scripted workflow with requirements that can be installed as a conda environment contained in an `environment.yml` file. The workflow is fully scripted and configurable with a `config.yml` file, and is run in 3 steps, with statistical analysis and visualization as an optional fourth step. First, a segmentation MRC file is converted automatically to a series of surface meshes formatted in the VTP file format. Second, for each mesh, the surface is converted to a graph (tg format) and curvature is estimated using pycurv. Third, orientations and distances between and within surfaces are calculated using the resulting graphs, and a CSV with quantifications as well as a final VTP surface file is output with all quantifications built in. Fourth, the outputs from multiple tomograms are combined for visualization and statistical analysis. Times and computational requirements are shown in supplementary table 1."

How this was addressed in the revised manuscript: These changes are included in the revised submission.

5. Figures 3-7 contain colorful 3D renderings of the measured quantities. In addition, they are filled with histograms of every possible quantitative parameter, which often are not very significant or different between. The authors should focus the main results and the figures to show the most relevant and significant findings and put the remaining panels and results into the supplement.

Proposed Revision Plan: Figures 3-7 were organized around the different methodologies (inter and intra-membrane spacing, curvature, orientation) but we agree that focusing to the main results of each methodology is sufficient to show the value of these results. We propose to address this criticism by moving figure 4D,F (inter-crista and junction spacing), figure 6 E,G (the junction measurements) and Figure 7 to supplemental figures. These supplemental figures will

also be joined by the previously requested OMM curvature analysis and our proposed analysis of autophagosomes.

How this was addressed in the revised manuscript: After separating surfaces and running our analyses on a per mitochondrion basis, we moved some panels from Figures 3-6 to Supplemental Figure 5. We also deposited the full dataset of per mitochondrion analyses for each measurement in Zenodo (10.5281/zenodo.7102596). We noted this in page 38, lines 906-907:

"*The complete output of the per mitochondrion analyses for every measurement is available at 10.5281/zenodo.7102596.*"

6. The exact morphological discrimination between fragmented and elongated mitochondria is not easily understood from the results section. What is really meant by blinded manual classification? It only became clear when reading the methods. The results section should stand on its own. How is the overall population between fragmented and elongated cells is affected after Tg application?

Proposed Revision Plan: To clarify our methodology for blinded classification of mitochondrial network morphologies we included the following text:

"*We categorized cells for mitochondrial network morphology by blinded manual classification in which five researchers were given fluorescence microscopy images of exemplar network morphologies (elongated and fragmented) as references to assign morphologies to the experimental fluorescence micrographs."*

We targeted similar ratios of elongated and fragmented cells in both vehicle and Tg treated conditions for tomography, but qualitatively saw the expected increase in the elongated population to what has been previously described during Tg treatment. Because of our single cell targeting approach we did not quantify the population shift."

How this was addressed in the revised manuscript: These changes are included in the revised submission.

7. Similarly, what is meant by manual classification of IMM, OMM and ER? Is there any clustering involved?
Proposed Revision Plan: Our automated segmentation approach labels all membranes, and the separation of the IMM, OMM, and ER membranes is done by an expert user selecting and relabeling each membrane based on cellular context (e.g. IMM is inside of OMM and contains cristae). We have added the following text to clarify our methodology for manual classification of IMM, OMM, and ER:

"*This was followed by manual labeling of membranes into mitochondrial IMM and OMM and ER membrane based on cellular context, as well as manual cleanup of individual membrane segmentations using AMIRA software (Thermo Fisher Scientific)."*

How this was addressed in the revised manuscript: These changes are included in the revised submission.

8. One of the key steps is the generation of a smooth surface from a segmented membrane, there is a question whether true membrane disruptions will be smoothed and may be overlooked in this approach. When these disruptions present true membrane ruptures, they may be of particular

biological importance. The authors should support the choice and selection of the smoothing parameters in order to illustrate this potential pitfall.

Proposed Revision Plan: The smoothing and hole-filling parameters are now configurable using the point_weight and extrapolation_voxels parameters in the config.yml file. Notably, the surfaces used for quantification used minimal smoothing, and any triangles more than a single voxel away from the point cloud were deleted, in order to ensure that the quantifications were minimally impacted by "hallucinated" surfaces. Additionally, the following text was added to the methods section discussion surface reconstruction:

"*A surface mesh was calculated from the oriented point cloud using the screened Poisson algorithm[32], with a reconstruction depth of 9, an interpolation weight of 0.7, and a minimum number of samples of 1.5. These settings were chosen to maximize correspondence to the data, rather than smoothness. The resulting surface extended beyond the segmented region, so triangles more than 1 voxel away from the point cloud were deleted. Interpolation weight (point_weight) and the mask distance (extrapolation_voxel) are both configurable in the surface morphometrics pipeline if more aggressive smoothing and hole filling are desirable.*"

How this was addressed in the revised manuscript: These changes are included in the revised submission.

9. Throughout the manuscript, the authors mention statistical significance several times and one of the main aims of the study is perform statistical hypothesis testing. It is important to specify the significance test (not only in the methods) and the p-value in order to support this claim. In the manuscript, the authors use exclusively the Mann-Whitney test. What is the rationale for choosing this test? Have the authors considered comparing the total distributions and not just the peaks with e.g. a Kolmogorov-Smirnov test? For a statistical methods paper, there are also no discussion on error analysis.

Proposed Revision Plan: This was a common concern raised by both reviewers, and we agree that a test based on total distribution would be more powerful than only looking at peaks. We address the use of the Kolmogorov-Smirnov test and the limitations we have run into thus far in our response to reviewer 1 in detail. In brief, KS tests tend to vastly overestimate statistical significance because the number of samples (the number of triangles) is vastly larger than the true number of independent features sampled in the data, so that even very similar looking distributions such as those in figure 5C yield p values in the range of 10^-200. We propose several approaches to better estimate the number of independent variables. We will also use a random subsampling approach within individual mitochondria to ensure sampling from the same distribution does not yield statistically significant results.

In addition to testing additional approaches to incorporate KS testing (based on estimation of number of independent features in each tomogram), we propose to improve our peak-based statistics by estimating a standard error for the peak of each tomogram using a bootstrap approach, getting the peaks from different random subsamples of triangles.

How this was addressed in the revised manuscript: In addition to the KS testing implementations we describe in response to similar comments from Reviewer #1, we incorporated a bootstrap sampling framework to assess the precision of our histogram peak measurements used for Mann Whitney U testing (and KS testing). The results of these bootstraps with 1000 resampling

iterations are given in Supplementary Table 3 and we added the following text to the results section on page 9 lines 231-235:

"*A challenge with using histogram peaks as a metric is determining the confidence interval associated with the metric. In order to assess the precision of the histogram peak measurements, we resampled each distribution using bootstrap sampling 1000 times and the histogram peak was extracted on each resampling. The 5th and 95th percentiles of these histogram peaks were used to represent the confidence interval of the histogram peak (Table S3).*"

Reviewer #2, minor comments:

1. https://github.com/grotjahnlab/surface_morphometrics should include an example data set or tutorial for dissemination.

Proposed Revision Plan: We are in the process of uploading all frame-averaged tilt series, tomograms, segmentations, and reconstructed surfaces to EMPIAR. Additionally, we propose to implement a complete tutorial including a single tomogram for readier workflow testing, separate from the complete data upload.

How this was addressed in the revised manuscript: As part of the v0.2 release of the surface morphometrics pipeline we added a pair of sample segmentations as well as a pre-configured `config.yml` file to the github repository for the surface morphometrics pipeline. We also include a tutorial describing approximate time information to perform processing of this test data on a midrange laptop.

2. What is meant by growth plane? This term is not defined in the manuscript.
Proposed Revision Plan: We clarified and further defined the "cell's growth plane" in the text by adding the following phrase:

"… the cell's growth plane (the plane of electron microscopy grid substrate on which the cell is grown) (Figure 6A)."

How this was addressed in the revised manuscript: This change is included in the revised manuscript.

4. What is meant by vehicle treatment? There is no explanation in the main text of the manuscript.

Proposed Revision Plan: We clarified and further defined vehicle treatment in the main text by adding the following:

"*We applied our correlative approach to identify and target specific Tg-treated and vehicle (media with DMSO) treated MEF^{mtGFP} cells with either elongated or fragmented mitochondrial network morphologies for cryo-FIB milling and cryo-ET data acquisition and reconstruction.*"

How this was addressed in the revised manuscript: This change is included in the revised manuscript.

5. Have the authors noticed/calculated any differences in the width of the cristae?

How this was addressed in the revised manuscript: We measure this difference in figure 4C (Intra-crista distance). We found significant changes in width/intra-crista distance in response to Tg treatment in elongated mitochondrial morphologies, and differences in the width/intra-cristae distance between elongated and fragmented morphologies in the Tg-treated conditions.

6. Methods: Automated surface reconstruction: “In cases where the resulting surface was very complex, the surface was simplified”…" How was the complexity determined?
Proposed Revision Plan: With the updated state of the software, we simplify all surfaces to generate a maximum of 150,000 triangles. This has minimal effect on very small surfaces, but greatly speeds computation on very large surfaces. We corrected the language to match this:

“*The resulting mesh was simplified with quadric edge collapse decimation to produce a surface that represented the membrane with 150,000 triangles or fewer*.”

How this was addressed in the revised manuscript: This change is included in the revised manuscript.

7. Methods: Calculation of distances between individual surfaces: “For surfaces with small numbers of triangles, this was accomplished using a distance matrix”…". What is the threshold for a small number of triangles?

Proposed Revision Plan: As part of our software overhaul we have changed to always using a more memory-efficient KD tree based quantification, since the additional speed for the distance matrix approach is minimal when there are few enough triangles for it to be appropriate, and the hardwired cutoff was not as flexible for different hardware configurations. The updated text is below, but to satisfy any potential reviewer curiosity, the decision was made when the required distance matrix would use more than 128GB of memory. In the case of two identically sized surfaces, this crossover happens when there are approximately 45,000 triangles in each surface.

“*For calculations of distances between respective surface meshes, the minimum distance from each triangle on one surface to the nearest triangle on the other surface was calculated using a KD-tree.”*

How this was addressed in the revised manuscript: This change is included in the revised manuscript.

Reviewer #2 (Significance (Required)):

The aim of the paper is well motivated. Cryo-ET is a growth field and there is a need for quantitative parameterization of cryo-ET data. Recently a toolkit for the analysis of filaments from cryo-ET has been published (Dimchev et al. 2021 DOI: 10.1016/j.jsb.2021.107808). Given the specific nature of the implementation, i.e. the membrane structures of mitochondria, I cannot easily see that this implementation will be useful beyond the analysis of mitochondrial membrane structure.

Proposed Revision Plan: We hope that we have addressed this concern with generality has been addressed by our previously described updates to the software implementation.

Revision Plan

5. **Description of analyses that authors prefer not to carry out.**

Review 2, minor comments:

4. Angle between OMM and cristae: Maybe use the average angle of each cristae for comparison or fit a plane for each cristae because you are interested in the angle between the cristae and the OMM and the membrane of the cristae has a lot of uneven surfaces

Proposed Revision Plan: We believe that the advantage of our approach is the ability to incorporate more complex geometric information from uneven surfaces such as those seen in cristae. With that said, the ability to quantify metrics for individual cristae in an automated manner would be very appealing, since in many ways cristae are functionally independent compartments. Accomplishing this would require either subdividing the larger surface into individual cristae, which will require development of additional sub-graph processing strategies. Additionally, pairing surfaces to represent opposite sides of a crista will require additional development. While we agree that this will be an excellent extension of the surface morphometrics approach, we feel that the additional development required is out of the scope of this initial manuscript focused on the general workflow. New methods leveraging sub-graph analysis will be explored in future manuscripts.

January 3, 2023

RE: JCB Manuscript #202204093R

Dr. Danielle Ann Grotjahn
Scripps Research Institute
Department of Integrative Structural and Computational Biology
10550 North Torrey Pines Rd
Hazen-173
La Jolla, California 92037

Dear Dr. Grotjahn,

Thank you for submitting your revised manuscript entitled "Quantifying organellar ultrastructure in cryo-electron tomography using a surface morphometrics pipeline." The manuscript was reviewed by original Reviewer #2 and new Reviewer #3. We would be happy to publish your paper in JCB pending the textual changes recommended by the reviewers as well as final revisions necessary to meet our formatting guidelines (see details below).

To avoid unnecessary delays in the acceptance and publication of your paper, please read the following information carefully.

A. MANUSCRIPT ORGANIZATION AND FORMATTING:

Full guidelines are available on our Instructions for Authors page, https://jcb.rupress.org/submission-guidelines#revised. **Submission of a paper that does not conform to JCB guidelines will delay the acceptance of your manuscript.**

1) Text limits: Character count for Tools is < 40,000, not including spaces. Count includes title page, abstract, introduction, results, discussion, and acknowledgments. Count does not include materials and methods, figure legends, references, tables, or supplemental legends.

2) Figure formatting: Tools may have up to 10 main text figures. Scale bars must be present on all microscopy images, including inset magnifications. Please avoid pairing red and green for images and graphs to ensure legibility for color-blind readers. If red and green are paired for images, please ensure that the particular red and green hues used in micrographs are distinctive with any of the colorblind types. If not, please modify colors accordingly or provide separate images of the individual channels.

3) Statistical analysis: Error bars on graphic representations of numerical data must be clearly described in the figure legend. The number of independent data points (n) represented in a graph must be indicated in the legend. Please, indicate whether 'n' refers to technical or biological replicates (i.e. number of analyzed cells, samples or animals, number of independent experiments). If independent experiments with multiple biological replicates have been performed, we recommend using distribution-reproducibility SuperPlots (please see Lord et al., JCB 2020) to better display the distribution of the entire dataset, and report statistics (such as means, error bars, and P values) that address the reproducibility of the findings.

Statistical methods should be explained in full in the materials and methods. For figures presenting pooled data the statistical measure should be defined in the figure legends. Please also be sure to indicate the statistical tests used in each of your experiments (both in the figure legend itself and in a separate methods section) as well as the parameters of the test (for example, if you ran a t-test, please indicate if it was one- or two-sided, etc.). Also, if you used parametric tests, please indicate if the data distribution was tested for normality (and if so, how). If not, you must state something to the effect that "Data distribution was assumed to be normal but this was not formally tested."

4) Materials and methods: Should be comprehensive and not simply reference a previous publication for details on how an experiment was performed. Please provide full descriptions (at least in brief) in the text for readers who may not have access to referenced manuscripts. The text should not refer to methods "...as previously described."

5) For all cell lines, vectors, constructs/cDNAs, etc. - all genetic material: please include database / vendor ID (e.g., Addgene, ATCC, etc.) or if unavailable, please briefly describe their basic genetic features, even if described in other published work or gifted to you by other investigators (and provide references where appropriate).

6) Microscope image acquisition: The following information must be provided about the acquisition and processing of images:
a. Make and model of microscope
b. Type, magnification, and numerical aperture of the objective lenses
c. Temperature
d. Imaging medium

e. Fluorochromes
f. Camera make and model
g. Acquisition software
h. Any software used for image processing subsequent to data acquisition. Please include details and types of operations involved (e.g., type of deconvolution, 3D reconstitutions, surface or volume rendering, gamma adjustments, etc.).

7) References: There is no limit to the number of references cited in a manuscript. References should be cited parenthetically in the text by author and year of publication. Abbreviate the names of journals according to PubMed. JCB formatting does not allow for supplementary references, please remove this section and add any non-duplicate references to the main reference list.

8) Supplemental materials: Tools papers are generally allowed 5 supplemental figures and 10 videos. You currently exceed this limit but, in this case, we will be able to give you the extra space. Each figure should span a single page so please either condense or split up the multi-page Figures S1, 3, & 4. Since you only have 6 main figures if you wish you can also move some of this material into the main text.

Please note that tables, like figures, should be provided as individual, editable files. A summary of all supplemental material should appear at the end of the Materials and methods section. Please include one brief sentence per item.

9) Video legends: Should describe what is being shown, the cell type or tissue being viewed (including relevant cell treatments, concentration and duration, or transfection), the imaging method (e.g., time-lapse epifluorescence microscopy), what each color represents, how often frames were collected, the frames/second display rate, and the number of any figure that has related video stills or images.

10) eTOC summary: A ~40-50 word summary that describes the context and significance of the findings for a general readership should be included on the title page. The statement should be written in the present tense and refer to the work in the third person. It should begin with "First author name(s) et al..." to match our preferred style.

11) Conflict of interest statement: JCB requires inclusion of a statement in the acknowledgements regarding competing financial interests. If no competing financial interests exist, please include the following statement: "The authors declare no competing financial interests." If competing interests are declared, please follow your statement of these competing interests with the following statement: "The authors declare no further competing financial interests."

12) A separate author contribution section is required following the Acknowledgments in all research manuscripts. All authors should be mentioned and designated by their first and middle initials and full surnames. We encourage use of the CRediT nomenclature (https://casrai.org/credit/).

13) ORCID IDs: ORCID IDs are unique identifiers allowing researchers to create a record of their various scholarly contributions in a single place. At resubmission of your final files, please consider providing an ORCID ID for as many contributing authors as possible.

14) Materials and data sharing:
As a condition of publication, authors must make protocols and unique materials (including, but not limited to, cloned DNAs; antibodies; bacterial, animal, or plant cells; and viruses) described in our published articles freely available upon request by researchers, who may use them in their own laboratory only. All materials must be made available on request and without undue delay. We strongly encourage to deposit all the cell lines/strains and reagents generated in this study in public repositories.

All datasets included in the manuscript must be available from the date of online publication, and the source code for all custom computational methods, apart from commercial software programs, must be made available either in a publicly available database or as supplemental materials hosted on the journal website. Numerous resources exist for data storage and sharing (see Data Deposition: https://rupress.org/jcb/pages/data-deposition), and you should choose the most appropriate venue based on your data type and/or community standard. If no appropriate specific database exists, please deposit your data to an appropriate publicly available database. Please, deposit your electron microscopy and mass spectrometry data in appropriate public databases.

B. FINAL FILES:

Please upload the following materials to our online submission system. These items are required prior to acceptance. If you have any questions, contact JCB's Managing Editor, Lindsey Hollander (lhollander@rockefeller.edu).

-- An editable version of the final text (.DOC or .DOCX) is needed for copyediting (no PDFs).

-- High-resolution figure and MP4 video files: See our detailed guidelines for preparing your production-ready images, https://jcb.rupress.org/fig-vid-guidelines.

-- Cover images: If you have any striking images related to this story, we would be happy to consider them for inclusion on the journal cover. Submitted images may also be chosen for highlighting on the journal table of contents or JCB homepage carousel. Images should be uploaded as TIFF or EPS files and must be at least 300 dpi resolution.

It is JCB policy that if requested, original data images must be made available to the editors. Failure to provide original images upon request will result in unavoidable delays in publication. Please ensure that you have access to all original data images prior to final submission.

The license to publish form must be signed before your manuscript can be sent to production. A link to the electronic license to publish form will be sent to the corresponding author only. Please take a moment to check your funder requirements before choosing the appropriate license.

Additionally, JCB encourages authors to submit a short video summary of their work. These videos are intended to convey the main messages of the study to a non-specialist, scientific audience. Think of them as an extended version of your abstract, or a short poster presentation. We encourage first authors to present the results to increase their visibility. The videos will be shared on social media to promote your work. For more detailed guidelines and tips on preparing your video, please visit https://rupress.org/jcb/pages/submission-guidelines#videoSummaries.

Thank you for your attention to these final processing requirements. Please revise and format the manuscript and upload materials within 7 days. If complications arising from measures taken to prevent the spread of COVID-19 will prevent you from meeting this deadline (e.g. if you cannot retrieve necessary files from your laboratory, etc.), please let us know and we can work with you to determine a suitable revision period.

Please contact the journal office with any questions, cellbio@rockefeller.edu or call (212) 327-8588.

Thank you for this interesting contribution, we look forward to publishing your paper in Journal of Cell Biology.

Sincerely,

Richard Youle, PhD
Monitoring Editor
Journal of Cell Biology

Dan Simon, PhD
Scientific Editor
Journal of Cell Biology

Reviewer #2 (Comments to the Authors (Required)):

The authors have taken the comments seriously and adapted the manuscript accordingly. In particular, the scope of the developed method revised manuscript is now significantly toned down and statistical significance has been addressed. The referee noted that the authors now correctly reference the underlying PyCurv tool. However, with the overhaul there are still a minor open item and a new item, which are not consistently presented.

1. The title of the manuscript has been adapted to pipeline rather than toolkit. However, the manuscript still uses the term toolkit a couple of times, e.g. in the abstract and at the short manuscript summary of the introduction (page 5).

2. The additional application of segmenting autophagosomal and ER membranes is interesting and puts the method on a wider foundation. However, technically speaking it is not clear whether the labeled presented structure really represents an autophagosome or lipid droplets or lysosomes. Unambiguous identification could only be shown with the appropriate CLEM labeled protein (Bieber et al.2022, PNAS). The respective reference detailed many different autophagosomal states and a multilammelar one is not amongst them. For the scope of this manuscript, it is not relevant but the pinpointing of this membrane structure to autophagosomes may be faulty and should be avoided. Showing the suitability of the method to multilammelar membrane structures should be completely sufficient for this manuscript.

3. In this context, as the autophagosomal membrane morphologies have been quantitatively analyzed in detail in the named Bieber reference, the authors should also cite that work as it also analyzes a multitude of membrane structures based on the previously developed PyCurv tools.

Reviewer #3 (Comments to the Authors (Required)):

This manuscript reports a pipeline for the analysis of membrane shapes in cryo-ET data, in particular mitochondria. It is probably the so far most systematic approach to an inherently difficult problem that many people in the field are facing; the quantification of pleomorphic, non-discrete ultrastructural features in cellular electron tomograms. The set of assessed features includes intermembrane distances, curvature, orientation of membranes relative to each other and assignment of membrane identity based on morphological features.

The here presented solutions will be extremely useful for the field, and certainly inspiring for further developments. Even if the presented results might not yet be the final solution to the overarching problem, they are crucial steps towards having more generally applicable and unbiased analysis workflows. I would therefore like to congratulate and thank the authors for their fantastic work!

On a technical level, the cryo-ET and cryo-CLEM work is of excellent quality, and the segmentation procedure is state-of-the-art. I am not really an expert in statistics, so I cannot judge all details regarding tests and data sampling, but as far as I can tell this is well done and convincing.

Nevertheless, there are a couple of issues that I would suggest to address:

- The conclusions are based on the classification into elongated vs fragmented mitos. I understand this classification is done by 5 researchers blindly. However, in an unperturbed cell population, I would expect a continuum between fragmented and elongated, rather than a completely binary population into two classes. Was this assessed? And further, it would be important to show the results of the 5-researcher classification, and assess how consistent the assignment of classes was between the 5 researchers.

- Related to the above: Fig 2A, the left panel is a bit confusing. It shows populations of cells that contain all types of morphologies, not only fragmented or only elongated mitochondria. The labels elongated and fragmented refer to only the cell that is highlighted by the dashed rectangle, is this correct? Please clarify in figure legend.

- Did Thapsigargin treatment change the ratio of elongated vs. fragmented mitochondria in the MEF population? If yes, changes in population distribution might have to be considered when conclusions are drawn from the statistical analysis of morphological features, which are compared in fragmented vs. elongated mitochondria. This is important throughout all analyses, as some of the observed Tg treatment effects might be due to a change in population size that skews the statistics.

- The section entitled "Bulk analysis and assessment of statistical significance based on surface quantifications" is very useful, because it provides a generalised approach to the analysis of morphological features, and it highlights some of the inherent difficulties of analysing such features. However it would be helpful to add a few introductory sentences to make clear what the purpose of this section is. While reading I was first confused that it did not specify what parameters where quantified and what are the peak shifts that are described. It took me several reads through the whole manuscripts to realise it.

- Also regarding that sections: On page 9, the compromise of 452 nm2 membrane area as the size of "independent" sample is explained. I understand the problem, and I agree this seems a good approach to tackle it. But it would be good to assess how robust the results are with respect to that exact number- would the significance of results change quickly if the feature radius was, say, 10 or 15 nm rather than 12 nm?

Minor:
- on page 10, "elongated cells" and "fragmented cells" is used, while it should say "cells with elongated / fragmented mitochondria".

- also on page 10: "decrease in the peak" or "increase in the peak" is a bit misleading, as what the authors mean is not an increase or decrease in amplitude but a shift on the x-axis. I recommend rephrasing.

- page 13/14: ER-mitochondrial contact sites: There is no reason to think that changes in contact sites are the cause of mitochondrial morphology changes. In fact, it could be the other way round. I suggest rephrasing to not confound cause and correlation.

- Legend to Supplemental Figure 7: Please spell out "autophagosome membrane", rather than "auto membrane".

- Suppl. Fig 9 mitochondria-ER contacts: in panel B, is the scale from 30 to 100 nm? The label 100 has a space between the second and third zero, which is confusing.
I also wonder if the label should not rather be <30 nm, as in panel C the measured distances are less than 30 nm.

Reviewer #2 (Comments to the Authors (Required)):

The authors have taken the comments seriously and adapted the manuscript accordingly. In particular, the scope of the developed method revised manuscript is now significantly toned down and statistical significance has been addressed. The referee noted that the authors now correctly reference the underlying PyCurv tool. However, with the overhaul there are still a minor open item and a new item, which are not consistently presented.

We appreciate the reviewers feedback and believe the overall scope of the manuscript was significantly improved by addressing their comments and concerns. We have addressed both the minor open item and new item in the revised manuscript, with the specific changes highlighted below.

1. The title of the manuscript has been adapted to pipeline rather than toolkit. However, the manuscript still uses the term toolkit a couple of times, e.g. in the abstract and at the short manuscript summary of the introduction (page 5).

We changed all instances of "toolkit" to "pipeline" in the manuscript text.

2. The additional application of segmenting autophagosomal and ER membranes is interesting and puts the method on a wider foundation. However, technically speaking it is not clear whether the labeled presented structure really represents an autophagosome or lipid droplets or lysosomes. Unambiguous identification could only be shown with the appropriate CLEM labeled protein (Bieber et al.2022, PNAS). The respective reference detailed many different autophagosomal states and a multilammelar one is not amongst them. For the scope of this manuscript, it is not relevant but the pinpointing of this membrane structure to autophagosomes may be faulty and should be avoided. Showing the suitability of the method to multilammelar membrane structures should be completely sufficient for this manuscript.

We appreciate the clarification and changed all instances of "autophagosomes" to "multilamellar membrane structures" in the manuscript text.

3. In this context, as the autophagosomal membrane morphologies have been quantitatively analyzed in detail in the named Bieber reference, the authors should also cite that work as it also analyzes a multitude of membrane structures based on the previously developed PyCurv tools.

We included a citation to the Bieber et al., PNAS 2022 paper and added in the following text on page 13 starting on line 374:

The observed multilamellar structures are comprised of multiple membranes[72], including the primary membrane (PM) and the secondary membrane (SM), as well as interior membranes (IM) (Supplemental Figure 9).

Reviewer #3 (Comments to the Authors (Required)):

This manuscript reports a pipeline for the analysis of membrane shapes in cryo-ET data, in particular mitochondria. It is probably the so far most systematic approach to an inherently difficult problem that many people in the field are facing; the quantification of pleomorphic, non-discrete ultrastructural features in cellular electron tomograms. The set of assessed features includes intermembrane distances, curvature, orientation of membranes relative to each other and assignment of membrane identity based on morphological features.

The here presented solutions will be extremely useful for the field, and certainly inspiring for further developments. Even if the presented results might not yet be the final solution to the overarching problem, they are crucial steps towards having more generally applicable and unbiased analysis workflows. I would therefore like to congratulate and thank the authors for their fantastic work!

On a technical level, the cryo-ET and cryo-CLEM work is of excellent quality, and the segmentation procedure is state-of-the-art. I am not really an expert in statistics, so I cannot judge all details regarding tests and data sampling, but as far as I can tell this is well done and convincing.

We appreciate the reviewer's enthusiasm about the impact and significance of our work!

Nevertheless, there are a couple of issues that I would suggest to address:

- The conclusions are based on the classification into elongated vs fragmented mitos. I understand this classification is done by 5 researchers blindly. However, in an unperturbed cell population, I would expect a continuum between fragmented and elongated, rather than a completely binary population into two classes. Was this assessed? And further, it would be important to show the results of the 5-researcher classification, and assess how consistent the assignment of classes was between the 5 researchers.

We agree with the reviewer that, within a given cell population, there is a continuum between fragmented and elongated mitochondrial network morphologies. For the scope of this manuscript, we elected to focus on the two most distinct and easily discernible network morphologies (i.e., elongated and fragmented) for our CLEM targeting. Given the relatively low throughput of CLEM/cryo-ET approach, we reasoned that focusing on two morphologies would enable us to increase the sample size for each morphology group, and would facilitate the development of our downstream quantitative and statistical analysis. In the future, we look forward to incorporating optimizations in throughput to the cryo-ET workflow that will enable us to generate larger sample sizes more quickly across a wide range of mitochondrial network morphologies and stress conditions!

- Related to the above: Fig 2A, the left panel is a bit confusing. It shows populations of cells that contain all types of morphologies, not only fragmented or only elongated mitochondria. The labels elongated and fragmented refer to only the cell that is highlighted by the dashed rectangle, is this correct? Please clarify in figure legend.

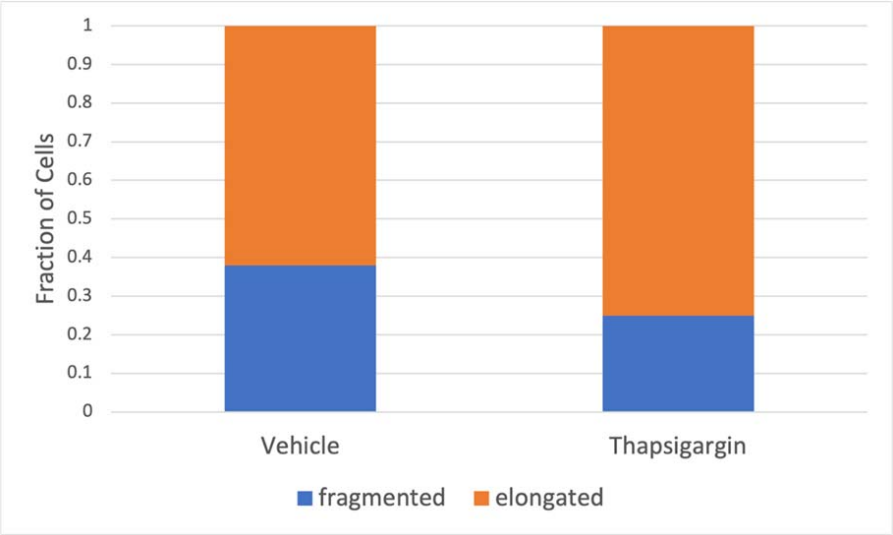
We appreciate the reviewer's concern and agree it is confusing as presented. To improve clarity, we moved the "elongated" and "fragmented" figure labels to the middle panel that displays a higher magnification view of the highlighted cell (to which the morphology label corresponds) and added the following text to the figure legend:

(A) Identification by cryo-FM of elongated (top; cyan outline) and fragmented (bottom; orange outline) mitochondrial morphologies from mixed populations of MEF^{mtGFP} cells. Cryo-FM images of bulk mitochondrial morphology are then used for targeted cryo-FIB milling to generate thin lamellae of ~150-200 nm. MEF^{mtGFP} cell periphery is outlined in dashed white line.

- Did Thapsigargin treatment change the ratio of elongated vs. fragmented mitochondria in the MEF population? If yes, changes in population distribution might have to be considered when conclusions are drawn from the statistical analysis of morphological features, which are compared in fragmented vs. elongated mitochondria. This is important throughout all analyses, as some of the observed Tg treatment effects might be due to a change in population size that skews the statistics.

We did observe a shift in the ratio of elongated vs fragmented mitochondria in MEF cells upon treatment with Thapsigargin towards a higher percentage of cells with elongated mitochondria (see figure below). This is consistent with previous reports (see Lebeau et al., 2018; PMID: 29539413). Given the relationship between mitochondrial inner membrane ultrastructure and function, we aimed to determine whether we could link these bulk, network morphology changes to quantifiable changes in mitochondrial membrane ultrastructure. While it is possible that the different distribution of fragmented and elongated mitochondria could impact the ultrastructural comparison between these two mitochondrial populations, our results suggest that, on a per cell basis, these two morphologies represent distinct functional states upon activation of ER stress (Tg-treatment). As we continue this project, we will integrate additional genetic and pharmacologic tools, in combination with our enabling CLEM/morphometric workflow, to gain further insight into the mechanisms defining the

relationship between mitochondrial morphology, ultrastructure, and function in the presence and absence of ER stress.



- The section entitled "Bulk analysis and assessment of statistical significance based on surface quantifications" is very useful, because it provides a generalised approach to the analysis of morphological features, and it highlights some of the inherent difficulties of analysing such features. However it would be helpful to add a few introductory sentences to make clear what the purpose of this section is. While reading I was first confused that it did not specify what parameters where quantified and what are the peak shifts that are described. It took me several reads through the whole manuscripts to realise it.

We added additional text for clarity and a more smooth transition between sections on page 8 starting on line 200:

Bulk analysis and assessment of statistical significance based on surface quantifications

Our pipeline outputs multiple membrane feature quantifications (i.e., curvature, distance, and orientation) for each triangle, totaling approximately 500,000 triangles per tomogram. We set out to determine an appropriate method for bulk quantitative and statistical analyses that would enable detection of subtle albeit potentially significant differences in membrane ultrastructure across different mitochondrial morphologies and treatment groups. We plotted the overall distributions of each membrane feature quantification using histograms of per-triangle values weighted by the area of each triangle, both for individual mitochondrion (Supplemental Figure 3) and for entire experiments. These histograms identify bulk differences in the membrane ultrastructure of mitochondria under different experimental conditions. For several membrane parameter quantifications, the observed histogram peak shifted visibly between conditions, suggesting potential differences between treatment and morphology groups. In order to test the statistical significance of these shifts, the weighted histogram peak for each membrane feature quantification for each individual mitochondrion was extracted and these peaks were used for statistical comparisons. In most cases, there was a statistically significant difference in the mean of the distributions, which was assessed via a non-parametric Mann-Whitney U test (Supplemental Table 2). Differences in the variability of metrics between mitochondria within a given experimental condition were assessed with a two-sample Kolmogorov-Smirnoff test.

- Also regarding that sections: On page 9, the compromise of 452 nm2 membrane area as the size of "independent" sample is explained. I understand the problem, and I agree this seems a good approach to tackle it. But it would be good to assess how robust the results are with respect to that exact number- would the significance of results change quickly if the feature radius was, say, 10 or 15 nm rather than 12 nm?

We appreciate the reviewer's concern and agree that it would be useful to test the robustness of this parameter on the results of the statistical output. As expected, when we run the statistics with varying feature radii (e.g.,

9nm and 15nm), the absolute p-values change, however the relative ordering of p-values among the different experimental groups does not change. Therefore, we believe the selection of 12 nm feature radius provides a reasonable, biologically-relevant metric for very roughly estimating the number of independent variables relevant to these experiments for the sake of making comparisons between conditions and statistics, although we do not argue that the absolute p-values are statistically sound. We added two additional columns with these values to Supplemental Table 2, and added the following lines of text to address on page 9 starting on line 256:

Altering the feature size to 9 or 15 nm changes the number of independent samples and absolute p values, however the relative p value differences across groups does not change, demonstrating that absolute value for assessing statistical significance is limited by the artificial selection of feature radius, but the choice of feature size does not affect the overall result of the comparative analyses across different membrane feature parameters (Supplemental Figure 2).

Minor:
- on page 10, "elongated cells" and "fragmented cells" is used, while it should say "cells with elongated / fragmented mitochondria".

We changed all instances of "elongated/fragmented cells" to "cells with elongated/fragmented mitochondria" in the manuscript text.

- also on page 10: "decrease in the peak" or "increase in the peak" is a bit misleading, as what the authors mean is not an increase or decrease in amplitude but a shift on the x-axis. I recommend rephrasing.

We appreciate the reviewer's clarification and we have now changed this section in the text on page 10 starting on line 272:

This revealed an overall increase in inter-membrane distance in fragmented mitochondria. Furthermore, we observed a subtle decrease in OMM-IMM distance after Tg-induced ER stress in elongated mitochondria, contrasted by a Tg-dependent increase in peak position and variability of distances in fragmented populations.

- page 13/14: ER-mitochondrial contact sites: There is no reason to think that changes in contact sites are the cause of mitochondrial morphology changes. In fact, it could be the other way round. I suggest rephrasing to not confound cause and correlation.

We agree this phrasing could be misleading and cause confusion. We changed this text on page 13-14 starting on line 384:

To investigate whether the remodeling we observed in response to Tg-induced ER stress was correlated with alterations to the ultrastructure of these contact sites, we leveraged our 3D morphometrics workflow to quantify changes in ER-mitochondrial contacts in elongated vs fragmented mitochondrial networks as well as in ER stress induced by Tg.
…
None of the conditions were statistically significant, suggesting that changes in ER-mitochondrial contact distance or area at the imaged time-point were not associated with the altered mitochondrial ultrastructure (Supplementary Figure 5C).

- Legend to Supplemental Figure 7: Please spell out "autophagosome membrane", rather than "auto membrane".Michaela to do

Per reviewer 1 suggestion, we changed all instances of "autophagosome" to "multilamellar membrane structures" and changed legend to supplemental figure 7 to "multilamellar membrane structure".

- Suppl. Fig 9 mitochondria-ER contacts: in panel B, is the scale from 30 to 100 nm? The label 100 has a space between the second and third zero, which is confusing.

I also wonder if the label should not rather be <30 nm, as in panel C the measured distances are less than 30 nm.

We reformatted the figure as recommended by the reviewer to increase clarity.