Reviewers' Comments:

First of all, we want to sincerely thank all the reviewers for their constructive and helpful criticism. We hope that we were able to address all of your concerns and suggestions with our substantial revision of the manuscript. Apart from considerable changes in the manuscript including light simulations to model aberrations, we extended the BigStitcher software with modules that allow for non-rigid registration, and virtual re-blocking to refine alignment, as well as a powerful method for quality estimation based on Fourier Ring Correlation.

*Reviewer #1:*

*Remarks to the Author:*

*Hörl et al. present a new software tool for automatically aligning large-scale 3D image tiles that were acquired using light-sheet microscopy including several refinement steps for optimal alignment. In addition to stitching tiles that were imaged from a single orientation, multiple pre-registered orientations can be fused as well, rendering the software suitable for very large specimens. The software is built upon several successful preexisting tools like the BigDataViewer, Bead-based Registration, ImageStitching and the MultiView Deconvolution algorithm by the same authors and implemented as a Fiji plugin, which will potentially make it useful for a large audience. The plugin is well-documented, includes sample data and should thus be easy to use for new users. The paper is nicely written, clearly structured and easy to follow. The figures are well designed and properly emphasize the possibilities of the framework. I have a few minor suggestions listed below but generally recommend accepting the manuscript.*

We thank the reviewer for this positive evaluation of our work.

*- The paper lacks a bit of discussion about the limitations of the framework. Maybe it would be good to extend the conclusions part a bit in this direction (only if there are any substantial limitations, of course).*

We added a section “Limitations” to the online methods/supplemental material. There is unfortunately not enough space for an in-depth discussion of these aspects in the main text itself. We therefore commented more on existing limitations in the respective paragraphs.

*- The new tool is based on several existing software packages and algorithms. It would thus be good to highlight, which of the parts are actually new algorithmic contributions and which parts were reused more or less unchanged from previous implementations.*

We highlighted the new developments of BigStitcher more in the text and additionally added significant functionality as outlined in the introduction (non-rigid, quality estimation, splitting).

*- Line 345 (main manuscript) and Supplementary Note on link dropping: here you refer to a distance threshold that is used to limit the correspondence assignments and a user-defined threshold for the average error. How does the user have to adjust these parameters? Is there any intuition on how to set them properly?*

We added the following text to the documentation, which hopefully makes it easier to understand and adjust:

The global optimization with iterative link dropping has two error thresholds:

* **relative error threshold:** the optimization will be repeated until the largest error of any tile is smaller than the average error of all tiles’ times the threshold. Lowering this threshold emphasizes consistent quality of the alignment.
* **absolute error threshold:** the optimization will be repeated until the average error of the tiles falls below this value, which should be set the average magnitude of displacements (in pixels) that is still "acceptable" after alignment.

**Both** stopping conditions have to be met for the optimization to finish.

Setting the thresholds can be a bit tricky, because both too high and too low values might lead to undesired results. Intuitively, one would set the 'absolute error' as low as is still acceptable and the 'relative error' rather low to ensure consistent quality. Setting the thresholds too low, however, will result in links being dropped until only a spanning tree of the link graph remains, in which case there is an "error-less" solution for placing the tiles, but it might be sub-optimal due to the emphasis on the remaining links.

Since the processing time for the global optimization is comparatively short and results are immediately displayed in BigDataViewer, manual optimization of these meta-parameters should be feasible, either starting with high values and lowering them or starting with low values and increasing them until the best and most consistent alignment is produced. We found, however, that the default values of 3.5px absolute error and 2.5 relative error work well in many cases. To give the user a bit more flexibility without confusing them with setting the values directly, we recently added "STRICT" (default thresholds) and "RELAXED" (default thresholds \*2) presets.

The documentation for the global optimization can be found at<https://imagej.net/BigStitcher_Global_optimization#Optimization_strategy_and_convergence_criteria>

*- The same applies to the regularization parameters \lambda\_1 - \lambda\_3 in Supplementary Note 7: how are they set properly? Any intuition or a rule of thumb how to specify them?*

We added a usage guide to the documentation, which can be found at <https://imagej.net/BigStitcher_BrightnessContrastAdjustment>

*- Figure 1: For panels (l, m) magenta is hardly visible. If this is in fact intended to show the perfect overlap, maybe mention that differences are hardly visible?*

We adjusted the images and the figure legend accordingly.

*- Generally, use consistent capitalization of the headlines (e.g., in the online methods there’s a mixture of title and sentence case)*

We fixed this inconsistency.

*- Supplementary Figure 1: Not sure what the benefit of showing the random forest classifications is to this figure. In my opinion, the main point the figure should emphasize is the preservation of fluorescence, i.e., a close-up of a few frames of the raw signal would be sufficient (potentially increase the zoom, such that structures can be easily resolved by eye).*

Our original rationale was that consistent detection by a random forest shows that the signal is preserved sufficiently, but we agree, it is not super intuitive. We therefore now show 6 slices with increasing depth of up to 2,0cm together with quality measurements based on the Fourier Ring Correlation to illustrate how deep the signal that can typically be detected.

*- Supplementary Figure 10: For DS2 and DS4 the distribution apparently becomes more narrow. Any idea why it broadens again for the DS8 plot?*

The reason is that downsampling smoothens the data, which, in combination with the subpixel localization of the spot, increases the registration quality at 2-fold and 4-fold downsampling. However, with even larger downsampling the loss of pixel resolution outweighs the effect of smoothing hence the quality drops. If, for example, you would have perfect data with no noise, the quality would drop already at 2-fold downsampling. We added this explanation to the online methods part.

*- Supplementary Figure 15: Interest points are hardly visible. Maybe show a magnification inset or generally increase the size of the interest point markers?*

We updated the figure and used screenshots made on a “normal” instead of a 4K display. The points are better visible in the real application, even on 4K screens. Now it should be nicely visible in the PDF.

*- Supplementary Figure 17: Scale bar is missing.*

We added the scalebars to the image.

*- It seems the test data is currently hosted on a GoogleDrive. Maybe it would be good to move the data to a different location, where it is also guaranteed to have persistent links that are also available in the future.*

We additionally added the example files to the Open Science Framework as a project: <https://osf.io/bufza/> It is one of the recommended websites by Nature (<https://www.nature.com/sdata/policies/repositories>). We also added the link to the supplement.

*- Supplementary Information Typos: know shift -> known shift, sice -> since, groping -> grouping, resonable -> reasonable, stong -> strong*

Thanks for pointing out those typos, we corrected them.

*- Supplementary Videos 2 and 5 are a bit too fast/short. I would recommend lengthening the videos, decreasing the playback speed a little or adding still frames in between, which would make it more easy to follow without having to restart the video multiple times.*

We updated both videos accordingly.

*Reviewer #2:*

*Remarks to the Author:*

*The article by Hoerl et al. describes an improved methodology, BigStitcher, to reconstruct large, modern fluorescence imaging experiments. The authors claim that BigStitcher is easier and more efficient to use than existing pipelines, automatically selects the best illumination direction, efficiently fuses and deconvolves multiview data, and compensates for all major optical effects in light sheet imaging.*

*BigStitcher is a much-needed advancement in the field, as many laboratories are now generating large imaging data sets without a rigorous pipeline to assemble, visualize, and analyze these data. While BigStitcher does not deal with analysis, the principled approach detailed in this manuscript for data assembly and visualizing will aid analysis and rigor in reporting by standardizing the first two steps in a pipeline.*

*In general, this is a well-written manuscript with clear experiments, results, and justifications. However, I do not believe the authors claim of compensating for all major optical effects is merited.*

*Unfortunately, the authors cannot control how users prepare their cleared tissue samples and carry out light sheet (or other high-throughput) imaging. These choices may render it impossible for BigStitcher to compensate for all major optical effects. My comments and questions are almost entirely concerned with this claim. I have attempted to provide a set of questions with actionable items for each major comment.*

We first of all we thank the reviewer for generally positive remarks on our software. We are very sorry that we created a confusion about what we actually intended to describe with “all major optical effects”. We agree with the reviewer that this is actually not the case, I think it mostly happened because of the space limit in the abstract. What we intend to claim is that we compensate for all major geometric transformations introduced by optical effects and also overall intensity adjustments. We changed the abstract and text accordingly. To further strengthen this point we added in-depth simulations of light propagation in tissues, added non-rigid capabilities, and implemented strategies to at least be able to robustly detect areas with reduced image resolution. We hope that this will address the concerns of the reviewer.

*Major comments and associated question for the authors:*

*1. Recent work by the Tomer group (Migliori et al. BMC Biology 2018) demonstrated why the exciting light sheet in a flat dual-sided light sheet configuration might not fully penetrate thick cleared samples. Knowledge of this optical aberration motivates the author's implementation of algorithms to decide which illumination direction provides the highest signal-to-noise. However, it is possible that one or both exciting light sheets will completely diverge while traversing a thick cleared tissue sample before reaching the imaging area.This situation will lead to a lack of usable information at that particular imaging plane, making it impossible to computationally correct for the responsible aberration.*

We agree and we see these types of aberrations in samples (Fig. 2a). We actually added simulations of light propagation that recapitulate this behavior and illustrate that we can now correct for it using affine, split-affine or non-rigid -- as long as the lightsheets stay in the depth-of-field of the common detection objective. We make it very clear that if that is not the case, the data cannot be reconstructed using BigStitcher: “… however, once the lightsheet is out-of-focus, blurred image data is acquired that cannot be reconstructed using BigStitcher.”

*If one light-sheet provides minimal information while the other complete diverges deep within a sample, there will be a high mean intensity signal from the diverged light sheet. The illumination detection schemes implemented in BigStitcher would incorrectly assign this high mean intensity signal as the 'correct' channel.*

*This false-positive identification is well known in autofocusing algorithms. An excellent discussion of the various options for autofocusing in light sheet fluorescence microscopy is found in Royer et al. Nature Biotechnology 2016. An application of a Fourier-based algorithm for autofocusing in cleared tissue imaging is found in Ryan et al. Nature Communications 2017.*

*While these manuscripts are mainly concerned with correcting first order defocus aberration during imaging, the computational principles of identifying the highest signal-to-noise carry over to selecting the best illumination direction.*

*Have the authors considered frequency-domain based algorithms to determine the best illumination pathway? How do these compare to the summary statistic based algorithms currently implemented in BigStitcher?*

Thanks a lot for this detailed input. We added additional methods (mean gradient & relative Fourier Ring Correlation) to determine the correct illumination direction and quantified their performance (Suppl. Fig. 2). Indeed, both perform much better than simple intensity measurements. We added some further description of the methods to the user guide: <https://imagej.net/BigStitcher_Select_illumination>

In general, our experience is that all three methods arrive at similar illumination selection results, with gradient magnitude and FRC providing more differentiation power in close cases (at the cost of longer compute times).

Motivated by another one of your comments, we additionally extended our relative FRC to enable image quality estimation throughout the sample, which turns out to be a

We also added the possibility to select illumination directions based on the frequency-domain based FRC, in addition to the existing methods (mean intensity and mean gradient magnitude).

The tool for measuring quality inside the specimen will be based on frequency domain measurements. IMPLEMENATION

*2. First order defocusin the imaging pathway is another common aberration in light sheet imaging. This aberration will lead to defocused images with minimal signal-to-noise. While the outcome is similar to the aberration discussed in Point #1, the cause of the aberration is different due to the orthogonal pathways in LSFM.*

*Because of the loss of signal-to-noise, the algorithms contained within BigStitcher cannot compensate for this aberration. The author's do not present a signal-to-noise plot as a function of depth from the surface of the cleared sample in this study. There is minimal information on experimental parameters (step size, confocal parameter, etc.) for the Zeiss Z.1 experiments as compared to the IsoView experiments.*

We added this plot (Supp. Fig 1)

Ok, add these parameters to the supplement. WRITING

I collected the information I have so far in a table

https://docs.google.com/spreadsheets/d/1CjofQ8D5v22Jovar7yn0W8i96xW-SGXTzCf-MZnmD94/edit?usp=sharing

*They also do not make clear if the Zeiss Z.1 or IsoView attempts to correct for first order defocus due to the refractive index mismatch between the imaging objective, refractive index matching media, and optically cleared or expanded tissue. This is particularly relevant for the larger sample imaged in the Zeiss Z.1.*

*It is not clear from the Figure 1, Figure 2, supplemental movies (Video 1 and Video 6) what the signal-to-noise deep within the cleared tissue sample is. Figure 2e is low-resolution imaging and box (iii) only shows the multi-view fusion of the data.*

Ok, we need to make clear that we work with whatever data we get. All we can do it alert the user if the quality is very low. WRITING

*I believe it is not feasible for the author to provide a link to the raw light sheet data because of the large size. Instead of this, more detailed discussion of the experimental parameters as well as improved supplemental figures and movies would demonstrate that BigStitcher is properly accounting for optical aberrations deep within the sample.*

We added Supp Fig 1 and FRC movies that illustrate the quality.

Ok, we need to make clear that we work with whatever data we get. All we can do it alert the user if the quality is very low. WRITING

*Have the authors performed a signal-to-noise analysis as a function of depth? For example, how does the Fourier Ring Correlation change as a function of depth? Can the authors please expand on the experimental details for the Zeiss Z.1 experiments?*

Again, supp Fig 1

Same as above let’s try to implement Fourier Ring Correlation as well for feedback to the user. IMPLEMENT & WRITE

*3. Higher order aberrations can be dealt with using deconvolution. Because of the depth-dependence of optical aberrations in cleared tissue, this can be a complicated and difficult task. The authors attempt to solve this problem by imaging diffraction limited fluorescent beads embedded within a hydrogel matrix. They do not present quantitative evidence that this approach is a feasible method for deconvolution. The obtained PSF are not provided, deconvolution results from image planes at the surface and at depth for each detection pathway are not provided, and deconvolution results from image planes at the surface and at depth for multiview deconvolution are not provided.*

*If the author's wish to claim that BigStitcher accounts for all major optical effects, theymust present a quantitative evaluation of their deconvolution scheme for thick cleared tissue.*

We do not wish to claim that. But we will show the PSFs and show some deconvolution results from different depths. ANALYSIS & WRITE

*4. Matsuda et al. Scientific Reports 2018 demonstrated fiducial free chromatic aberration correction. A key point of their work is that the same features must be imaged in all detection channels using one excitation color for accurate correction if using interest points. On Line 108 of this manuscript the author's state that their approach works if 'autofluorescence levels are sufficiently high.' This seems to be a critical point and should be reinforced to users that interest point extraction followed by ICP cannot correct for all chromatic aberration unless the same structure is present through all channels when excited with the lowest wavelength excitation (e.g., DAPI across all images without other excitations).*

We will make this clearer in the manuscript. WRITE

*Have the authors quantified how this single-excitation/multiple-detection chromatic correction scheme compares with their current interestpoint across all channels method?*

I am not sure what the question is here. READ & ACQUIRE & WRITE>> FABIO

What they do is that they have a software where you can acquire a reference image in a channel like 405 and by detecting the bleed through in the other channels you can correct for the chromatic aberration by calculating translation, rotation, magnification. Then you apply this model to your target image. They test they method on simulated data and achieve a deviation from the truth after correction around 16 nm

Here is a quick guide of what they do:

<https://github.com/macronucleus/Chromagnon/releases/download/Doc-v0.5/DocumentV064.pdf>

So Matsuda et al uses the single-excitation/multiple detection and we use Interest Points from different excitation-emission channels.

I don’t know exactly what the reviewer wants. One option is to compare the chromagnon chromatic shift (single-excitation/multiple-detection) to our chromatic correction (IP across all channels) or acquire an image single-excitation/multiple-detection) and an image single-excitation/single-detection) and compare the correction of the chromatic shift using our IP method

*Minor comments:*

*1. We attempted to evaluate BigStitcher on tiled 3D stacks with 10 x 10 XY tiles with 20% overlap, 51 z positions, 4 colors, 2 spectrally separated detection sCMOS cameras (OrcaFlash4.0 v2) each with 2048x2048 pixels. These stacks were acquired using the Multi-Dimensional Acquisition plugin in MicroManager 2.0 gamma build 2018.04.01. BigStitcher was unable to parse the metadata correctly for tile positions outputted from MM 2.0. We had to arrange the tiles using the built-in functions manually.*

*It would be helpful to have a 'how-to' guide on parsing the metadata from the MDA plugin into a BigStitcher compatible format so that the stage positions could be utilized.*

We added advice how to do another way via TIFF.

Explain that one should open a ticket for a special workflow. Problem is that Bioformats does not seem to read MM files correctly. If we had such a dataset that fails we could easily correct that, show past fixes. Still, it works, which is great. WRITE

We tried to reproduce the error using simulated hardware (DemoStage, DemoXYStage, 2 DemoCameras) in micro manager (same build, macOSX) as follows:

\* We set up a MDA with multiple, hand-picked, XY positions, relative z stacks and 2 channels, each using a different DemoCamera

\* We chose to save the Images as Image stack files: MMStack\_Pos\*.ome.tif

We then imported the data into BigStitcher as follows:

\* Define a new dataset using the Automatic Loader (Bioformats based)

\* As the file path, we used /path/to/images/MMStack\_Pos\*.ome.tif

\* We mapped BioFormats Series to Tiles (BioFormats Series are? -> Tiles)

\* For the numerical pattern detected in the files (MMStack\_Pos{0}.ome.tif) we chose to --ignore this pattern-- (in the newest version, this pattern is ignored automatically)

\* Under Move to Grid, we chose “Do not move Tiles to Grid (use Metadata if available)

By following this procedure, BigStitcher seems to be able to load xy positions from metadata correctly, at least for the DemoXYStage.

If you tried the same steps on your setup an it did not work, we would of course be happy to look at your data and figure out the problem, if you could provide a small example dataset anonymously somehow. In general, we are happy to receive feature requests and bug reports via <https://github.com/PreibischLab/BigStitcher/issues>.

Other users have reported problems loading MM data as well (<https://github.com/PreibischLab/BigStitcher/issues/37>), so we fixed the problem of having to select the counterintuitive “--ignore this pattern--” and we provide an easy way of flipping images if your stage coordinates do not comply to our x=left, y=down scheme.

We’ve also recorded a screencast showing how to import TIFF datasets (including MM): <https://youtu.be/aUofNP6V0lg>

*2. On this same dataset (and four replications of this dataset from different preparations), we attempted to use the interest point extraction followed by ICP. Each channel is 2048x2048 pixels with 51 z positions separated by 250 nm. An Olympus 40x NA 1.3 oil immersion objective was utilized for excitation and detection for an effective pixel size of ~162 microns. We intentionally introduced a slight mismatch in XY position between the two spectrally separated detection cameras to test the chromatic aberration algorithm. An imaging flat beamsplitter from Semrock directed emitted photons due to 405/488 excitation to camera #1 and photons due to 555/640 excitation to camera #2. There is some spectral bleed-through of photons due to 488 excitation to camera #2.*

*The Register with ICP options 'Simple (tile registration)' or 'Simple (all together)' options causes the axial extent of the stitched image to expand by almost ten-fold with huge distortions in the channel associated with 405 nm excitation (DAPI nuclear counterstain in this case). However, only performing 'Simple (chromatic aberration)' correctly aligned the offset in the cameras with no axial distortion.*

*This behavior seems abnormal. Perhaps we are utilizing the software incorrectly?*

Add some DOCUMENTATION.

The ICP algorithm might produce these erroneous results if no reliable point matches can be determined. Point matches are iteratively determined by matching interest points in one image to their nearest neighbor in the other, calculating and applying a least-squares optimal transformation to align the point sets and repeating the process until it converges. In the presence of too few actually correct interest points present in both images (e.g. due to too many detections in the background) or a too large neighbor search radius, the procedure might converge to a matching of actually unrelated points, and thus a bad registration.

One requirement of ICP is that the images are already roughly aligned. For tile registration, it should therefore be used as a refinement step after an initial alignment with phase correlation. Did you by any chance perform the ICP before stitching via phase correlation (guess because your question about that comes after this)? In this case, the observed behavior would make sense, as the channels, with only a slight shift are sufficiently pre-aligned for ICP to work, while the tiles are not pre-aligned enough and erroneous  registrations are generated from the unreliable point matches.

Otherwise, playing with the detection parameters (downsampling and threshold) to detect more sensible interest points (they can be visualized in Multiview mode) or adjusting the neighbor search radius (Fine/Normal/Gross alignment) might help. Furthermore, by going to the Expert... Option for the ICP refinement, one could restrict the transformation model to rigid (no scaling/shearing), apply stronger regularization to the model or use manually detected InterestPoints (detected in Multiview mode) for the refinement.

We have in-depth documentation for this step of the processing pipeline on the ImageJ wiki:<https://imagej.net/BigStitcher_ICP_refinement>

*3. It is not clear how the 'best' n points are selected during the Phase Correlation step. Can the authors please clarify the process of the algorithm?*

Add this in the supplement and/or documentation. WRITE.

Ideally, the phase correlation algorithm produces a phase correlation matrix (PCM) with a single intensity peak, whose location corresponds to the shift vector between the two input images. However, in real images with noise and repetitive structures, the PCM might contain multiple peaks. Furthermore, due to the circular nature of the Fourier shift theorem, each peak might correspond to 2^number of dimensions actual shifts. In our implementation, we therefore check the n peaks (subpixel-localized local maxima) of highest intensity and validate each possible shift by calculating the cross correlation of the shifted images. The final result is the shift yielding the highest cross correlation.

The procedure is explained under Pairwise Stitching using Fourier-based Phase Correlation in the online methods as well as in the BigStitcher documentation on the ImageJ wiki: https://imagej.net/BigStitcher\_Advanced\_stitching#Advanced\_phase\_correlation

*4. How does the BigStitcher multiview fusion compare to the CUDA-based IsoView multiview fusion presented in Chhetri et al. Nature Methods 2015? A quantitative comparison with the instrument's original reconstruction algorithm would strengthen the author's argument that BigStitcher incorporates multiple views in a streamlined pipeline that corrects for major optical effects.*

“we wrote our own implementation of the Lucy-Richardson three-dimensional multiview deconvolution algorithm4,17,24”. It is a re-implementation of the same algorithm that is used here. Point out that the contribution here is that you can run it on a terabyte dataset, which the CUDA software cannot. WRITE(ask Philipp)

*5. We find the memory estimates that BigStitcher provides for image fusion are consistently incorrect. Using a workstation with 128 GB of RAM, we regularly run out of memory for stitching processing that the BigStitcher plugin estimates will only require ~90 GB. Providing accurate estimates to users will streamline their use of the plugin and aid adoption. Particularly because the plugin offers an alternate fusion strategy that does not require as much RAM.*

Let’s adjust that to be more precise. IMPLEMENT

*Reviewer #3:*

*Remarks to the Author:*

*This paper describes the BigStitcher software for reconstruction of large (TBs sized) multi-tile, multi-view light microscopy acquisitions of whole-tissue and whole-organism.*

*I enjoyed reading this paper. The paper is well written and the chain of thoughts are easy to follow. The results look very promising and useful to the community.*

*Can the authors please comment to why their tool is significantly slower in execution time than most of the other tools for the Fusion step (as shown in table 1)?*

Implement Translation-only fusion, IMPLEMENTATION

Explain difference to affine more clearly, link to simulation why it’s necessary WRITING

*I am surprised that the authors did not address the accuracy of their stitching method. Stitching is the first computational step when conducting a large scale biological experiment and some conclusions depend on the quality of the stitched image.*

*Do the authors intend to quantify the accuracy of their stitching tool in the future? Visual inspection is not enough especially for such large specimen.I would like to point the authors to a recently developed tool:* [*https://www.nature.com/articles/s41598-017-04567-y*](https://www.nature.com/articles/s41598-017-04567-y) *that does microscopy stitching in 2D where a reference dataset was created to quantify the stitching accuracy of the developed tool.*

Implement simulation, show how it works. IMPLEMENTATION

Use existing datasets as benchmark. BENCHMARKING