nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-019-0501-0

BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples

David Hörl[1,2,7], Fabio Rojas Rusak[3,7], Friedrich Preusser[2], Paul Tillberg[4], Nadine Randel[4], Raghav K. Chhetri[4], Albert Cardona[4,5], Philipp J. Keller[4], Hartmann Harz[1], Heinrich Leonhardt[1], Mathias Treier[3,6] and Stephan Preibisch[2*]

Light-sheet imaging of cleared and expanded samples creates terabyte-sized datasets that consist of many unaligned three-dimensional image tiles, which must be reconstructed before analysis. We developed the BigStitcher software to address this challenge. BigStitcher enables interactive visualization, fast and precise alignment, spatially resolved quality estimation, real-time fusion and deconvolution of dual-illumination, multitile, multiview datasets. The software also compensates for optical effects, thereby improving accuracy and enabling subsequent biological analysis.

Sample clearing[1] and expansion microscopy (ExM)[2] are powerful protocols that create large transparent volumes of whole tissues and organisms (Fig. 1a,b and Supplementary Notes 1–3; Methods). Using light-sheet microscopy[3–5] (Fig. 1c), these samples can be imaged with subcellular resolution in their entirety within a few hours[6]. These acquisitions have the potential to be powerful tools for whole-tissue and whole-organism studies because they preserve endogenous fluorescent proteins (Supplementary Fig. 1) and are compatible with most staining methods.

However, the raw data acquired with the microscope are not directly suitable for visualization and analysis. Many large overlapping three-dimensional (3D) image tiles are collected that amount to many terabytes in size and require image alignment (Fig. 1d–n). Owing to sample-induced refraction and scattering of the light sheet in the direction of illumination[7], 3D image tiles are typically acquired twice with alternating illumination from opposing directions[5] to achieve full coverage (Figs. 1d and 2, and Supplementary Fig. 2). Similarly, emitted light is distorted by the sample, effectively limiting the maximal imaging depth at which useful data can be collected (Fig. 1n and Supplementary Fig. 1). Additionally, overlapping images suffer from spherical and chromatic aberrations (Supplementary Figs. 3 and 4). For reconstruction, and to make these complex datasets easily accessible to biologists and computer scientists, we developed the BigStitcher software. It enables interactive visualization using BigDataViewer[8], fast and precise alignment, quality estimation, real-time fusion, deconvolution and alignment of multitile acquisitions taken from different physical orientations (so-called multitile 'views'), thereby effectively doubling the size of specimens that can be imaged (Fig. 1n), while further orthogonal views can render the resolution isotropic.

BigStitcher features a new user-friendly importer for a multitude of vendor-specific and custom formats that is based on BioFormats[9], and accesses image data through memory-cached virtual loading[10], which can optionally be combined with virtual flat-field correction (Supplementary Figs. 5 and 6, and Supplementary Notes 4 and 5). Performance is optimal when images are initially converted to a multiresolution, blocked and compressed format such as HDF5 (ref. [8]) enabling interactive visualization, processing and interaction with terabyte-sized image datasets.

Accurate reconstruction of these large complex datasets requires compensation for the different types of image and intensity transformations that are introduced by the acquisition process. We therefore developed an interactive stepwise process that compensates for all relevant transformations while providing spatially localized feedback on the quality of the acquired image data (Supplementary Figs. 7 and 8).

First, overlapping 3D image tiles are acquired to cover the entire sample for each acquisition angle. Although approximate tile locations are typically known, translation stages usually show substantial inaccuracies (Supplementary Fig. 9). To compute locations for every image tile, we developed an image stitching algorithm optimized for very large datasets that can deal with acquisitions arranged in non-regular grids (Fig. 3a) containing empty images and multiple independent samples (Supplementary Fig. 10). As acquisitions often consist of hundreds of gigabyte-sized image tiles, each containing very different information (Fig. 3), we initially compute each shift for all pairs of overlapping tiles (links) using the parameter-free phase correlation method[11–14] on downsampled images. It computes all possible shifts between two images, and intensity peaks in the resulting phase correlation matrix correspond to shifts with high correlation that we localize with subpixel precision (Supplementary Fig. 11 and Supplementary Note 6; Methods). Using simulations, we showed that our new pairwise stitching method achieved errors below 1 pixel while reducing computation times 100-fold (Supplementary Figs. 12–15 and Supplementary Note 7; Methods). As correlation-based approaches can fail for image pairs characterized by repetitive patterns, noise or low information content, computing final image tile locations requires global optimization, which sometimes needs to be combined with manual curation (Supplementary Fig. 16 and Supplementary Video 1). Our new global optimization method extends the concept of identifying tile positions by minimizing the distance between all links[12,15], which, in comparison to computing a minimum spanning tree[16], averages out normally distributed link errors (Supplementary Fig. 15) dur-

[1]Department of Biology II, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [3]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [6]Charité-Universitätsmedizin Berlin, Berlin, Germany. [7]These authors contributed equally: David Hörl, Fabio Rojas Rusak. *e-mail: stephan.preibisch@mdc-berlin.de

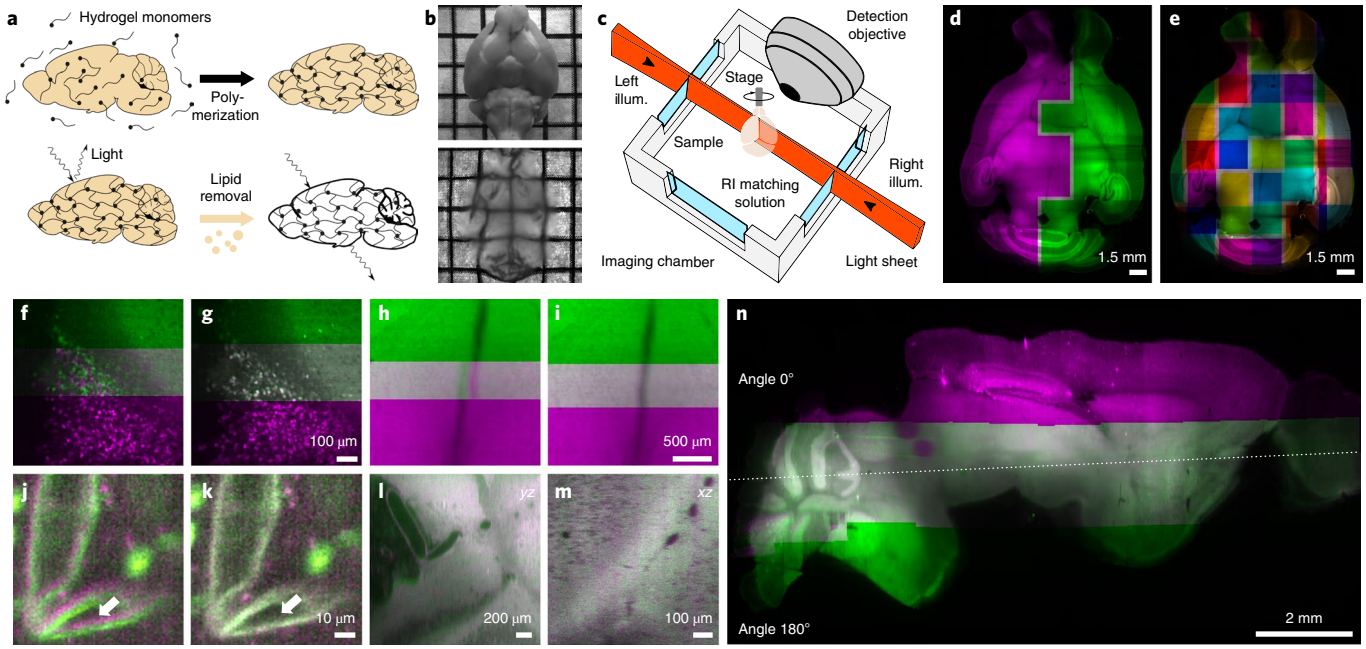

Fig. 1 | BigStitcher principles. a, Schematic of the CLARITY sample clearing process. **b**, Adult mouse brain before (top) and after (bottom) clearing. **c**, General layout of the type of light-sheet microscope used for most acquisitions[3,5]. RI, refractive index. **d**, A single slice through an entire adult mouse brain acquired with dual-sided illumination. The left illumination (illum.) is shown in pink and the right illumination is shown in green; image tiles indicate the illumination direction that was automatically selected for each tile. **e**, Overview of an entire section of an acquired adult mouse brain; different colors indicate individual image tiles (each 1920 × 1920 × 770 pixels). **f**–**i**, Illustration of the result of image stitching from a mouse brain expressing the histone 2B–eGFP lineage tracing marker in BSX-expressing neurons ($Bsx^{H2BeGFP}$) using phase correlation before (**f**,**h**) and after (**g**,**i**) stitching. **j**,**k**, The effect of ICP refinement on two different channels with sufficient autofluorescence visible in both channels. Arrows indicate the difference before (**j**) and after (**k**) refinement. **l**,**m**, Illustration of the high-quality multiview reconstruction for two overlapping multitile views at 0° (magenta) and 180° (green) for an axial versus axial (**l**) and lateral versus axial (**m**) view. **n**, One slice through an entire adult mouse brain (2.24 terabytes of raw data); both views are shown in axial orientation looking along the rotation axis of the microscope. The dotted line illustrates the middle of the section. In **b**,**d**–**n**, clearing and alignment (Methods) were performed on six independent samples with similar results (Supplementary Table 1).

ing optimization, as tiles are typically linked to many neighbors (Fig. 3a). Incorrect links are filtered by quality and by iteratively removing the link that disagrees most with the global optimization result[12] using a new compound metric. In current implementations, unconnected tiles (for example, empty images) and multiple independent objects in an acquisition are handled by ignoring them[12,13], or assuming regular two-dimensional (2D) translational grids[14]. Here we present a generic solution to this problem by introducing the concept of strong and weak links (Supplementary Fig. 10), which is independent of the original tile arrangement and not limited to translations. Strong links correspond to confirmed links, while weak links are derived from current transformations (for example, approximately known tile positions). Optimizing both link types in an acquisition yields accurate registration results within strongly linked regions and optimal alignments for weakly linked groups of tiles (Fig. 3a and Supplementary Fig. 10; Methods). However, correct tile placement (that is, solving the classical stitching problem) represents the first step and is usually not sufficient to properly align dual-illumination, multitile datasets (Fig. 2).

Second, because microscopy images suffer from spherical and chromatic aberrations that can be approximated reasonably well by affine transformations if distortions are in the range of a few pixels (Supplementary Figs. 3 and 4), we implemented a single-step interest-point-based alignment step that automatically extracts interest points and applies a variation of the iterative closest point algorithm (ICP)[17] on affine transformations. In combination with our new global optimization BigStitcher is able to compensate for small affine distortions that arise from spherical, and also chromatic aberrations if the same autofluorescent structures are visible across channels (Fig. 1i,j and Supplementary Figs. 3, 4 and 9; Methods).

Third, although samples are highly transparent (Fig. 1b), light scattering is an issue when imaging in tissues at depths on the order of centimeters. Although improved designs were recently proposed[18], dual-sided light-sheet illumination[5] (Fig. 1c,d) remains the most prominent method to double the sample size for which high-resolution image data can be collected laterally. Before alignment, we automatically suggest the best illumination direction for each tile by estimating image sharpness (Fig. 1d) using newly developed methods (Supplementary Fig. 2; Methods). Unexpectedly, we observed non-rigid image deformations between image tiles with different illumination direction (Figs. 1d and 2a–d). To understand how these transformations are created, we performed simulations of light propagation in tissue using ray tracing (Fig. 2e–h; Methods). These simulations showed that refractions within the illumination light path can lead to different parts of the sample being illuminated, and these parts can both lie in focus of the same detection objective owing to typical depths of field being in the range of several tens of micrometers (Supplementary Table 1). To compensate for these transformations, we implemented a virtual non-rigid alignment method based on identified corresponding interest points[19], as well as a piecewise ICP-based affine alignment based on virtual splitting of image tiles into smaller blocks (Fig. 2b–d,i–m; Methods). Depending on the magnitude of refraction, affine, split-affine or non-rigid alignment was the best choice for precise alignment (compare Fig. 2a–d and Fig. 2k–m), which is possible as long as the light-sheet remains within focus. However, once the light

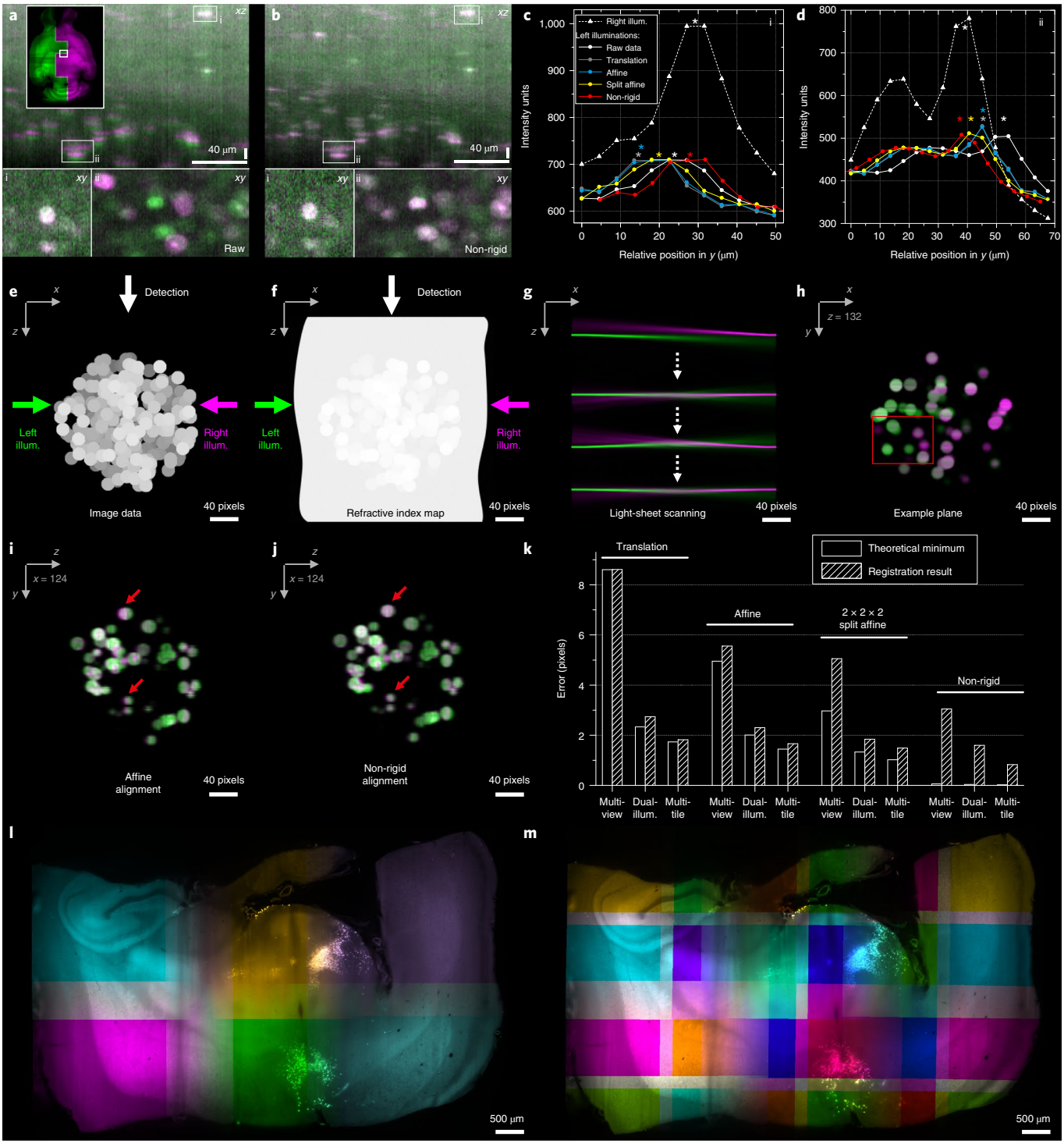

Fig. 2 | Optical aberrations, light simulations and alignment quantification. a, Expanded view of an area that shows misalignment between the two different illumination directions (left and right illuminations are shown in green and pink, respectively) of the original acquired images. An overview of the entire mouse brain is shown for orientation. Two selected areas (labeled i and ii) from the *xz* view are shown in the *xy* view below, highlighting the misalignment. **b**, The same views as in **a**, but corrected using non-rigid alignment; again, two are shown in the *xy* view below. **c**,**d**, Plots of cross-sections of the selected areas in **a** (ii plotted in **c**) and **b** (ii plotted in **d**) comparing different BigStitcher alignment methods. Asterisks mark the approximate centers of nuclei, which should ideally overlap between different illumination directions. **e**–**h**, Simulation of a spheroid-like object using ray tracing that recapitulates (red box) the non-rigid deformations indicated in **a**(ii). Supplementary Video 2 shows a corresponding animation. **e**, Simulated image data. **f**, Refractive index map. **g**, Illustration of the effects on the light sheet while scanning. **h**, An example slice of the simulated object. **i**,**j**, Comparison of the alignment quality for the affine and non-rigid alignments of the simulated data. **k**, Comparison of the best theoretically achievable registration quality from the manually selected corresponding points and a typical registration result when using different alignment modes in BigStitcher using the datasets shown in **l** and **m**. **l**,**m**, An as-small-as-possible 166-gigabyte multiview, dual-illumination, multitile dataset specifically acquired for error quantification in the original configuration (**l**) and after the virtual 2×2×2 split (**m**). In **a**–**m**, Non-rigid deformations of varying degree were observed for all dual-illumination datasets (Supplementary Table 1), a single simulated dataset was created and a single dataset for precise quantification was acquired. Different affine and non-rigid alignments were applied to those three datasets.

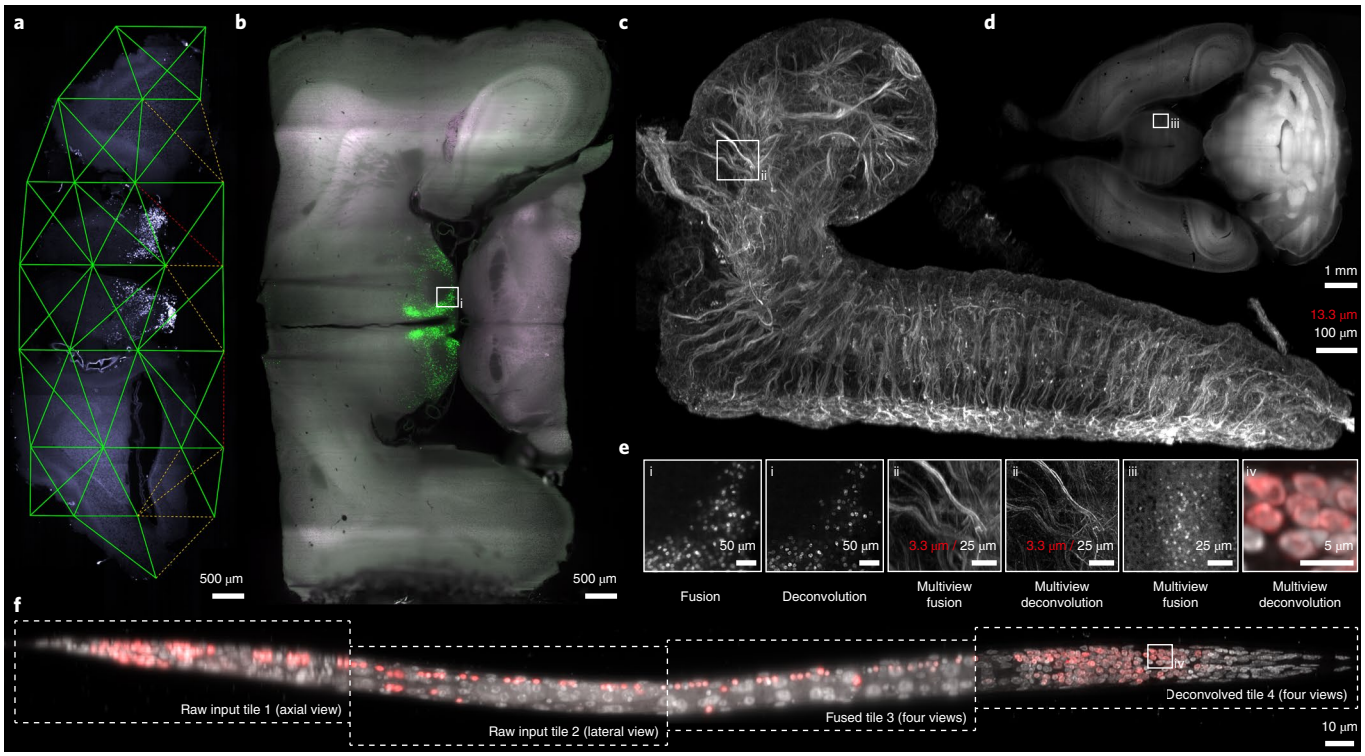

Fig. 3 | Reconstructed samples. a, One slice through an acquisition of an adult mouse $Bsx^{H2BeGFP}$ coronal slice encompassing the hypothalamus (Supplementary Video 3). Green lines indicate strong links between overlapping image tiles, dotted orange lines indicate links rejected because of low correlation and red lines indicate links that were determined to be unconcise. **b**, One slice through an adult $Bsx^{H2BeGFP}$ mouse brain. **c**, Maximum-intensity projection of the central part of the 7.5-fold expanded central nervous system from a *Drosophila* first instar larva with immunostaining for tubulin (Alexa Fluor 88) imaged with multitile IsoView light-sheet microscopy (Supplementary Videos 4 and 5), red indicates scale before expansion **d**, One slice through a whole multitile, multiview reconstructed adult $Bsx^{H2BeGFP/+}$ mouse brain (Supplementary Video 6). **e**, Expanded views of specific areas from **b**–**d** and **f** (labeled i–iv) illustrating (sub)cellular resolution and the advantage of (multiview) deconvolution over (multiview) fusion. **f**, A fixed *Caenorhabditis elegans* dauer larva acquired in four tiles with four views each; tagRFP is expressed in the nuclei of all neurons, which are co-stained with DAPI (Supplementary Video 7). Boxes indicate the quality of axial and lateral raw input data, multiview fusion and multiview deconvolution. In **a**–**f**, a total of eight different datasets were acquired for this publication (Supplementary Table 1), which were all reconstructed as described in the Methods. Additionally, the BigStitcher reconstruction pipeline has been applied to more than 50 samples in our lab (data not shown).

sheet is out of focus, blurred image data, which cannot be reconstructed using BigStitcher, are acquired. Such first-order defocusing can, however, be minimized by employing autofocusing during the acquisition process[20,21].

Finally, as emitted light is distorted by the sample, maximum imaging depth is limited. To overcome this problem, we acquired samples from opposing directions by rotation (Fig. 1c) or by simultaneous acquisition with two objectives[4]. We developed an optimized method for registration of large multitile views, where each view consists of a set of aligned image tiles from one physical orientation (Supplementary Note 8). This method robustly aligns large volumes using affine transformations, effectively doubling the imaging depth of any sample (Fig. 1n and Supplementary Table 2). Subsequently, applying ICP-based non-rigid, split-affine or affine registration allows precise multiview alignment that accounts for additional light refraction in the excitation light path. Using example data, we quantified theoretically possible and practically achievable registration performance (Fig. 2c,d,k, Supplementary Figs. 3, 4 and 17, and Supplementary Note 9), which illustrates that translations alone are not sufficient to achieve high-quality image reconstructions.

As image quality is not constant across the sample, it needs to be quantified to ensure that every part of the reconstructed dataset was acquired with high quality. However, manual inspection at the highest resolution for the entire sample is impossible owing to its size. Therefore, we developed relative Fourier ring correlation (rFRC), which is based on Fourier ring correlation (FRC)[22]. rFRC is able to automatically and rapidly estimate image quality throughout terabyte-sized light-sheet acquisitions while accounting for common scientific complementary metal oxide semiconductor (sCMOS) camera patterns (Supplementary Figs. 1, 7 and 8; Methods).

For downstream analysis, datasets can be fused or directly analyzed using BigDataViewer plugins. We implemented a new algorithm for real-time fusion by multithreaded processing of the currently visible plane in virtual images using blockwise multiresolution loading (Methods), which can optionally be performed with downsampling and on regions of the sample (Supplementary Fig. 18) while supporting brightness equalization. It enables fusion of terabyte-sized images on machines with little memory (Supplementary Fig. 19), while increased memory and compute power enable faster processing (Supplementary Table 2).

Deconvolution is an established method to increase contrast and resolution in light microscopy acquisitions, and required point spread functions (PSF) are typically estimated using fluorescent beads[23]. To handle multitile, multiview acquisitions, we extended deconvolution code[23] allowing BigStitcher to deconvolve selected regions and improve image quality (Fig. 3b–f and Supplementary Note 10). To estimate required PSFs in cleared samples, we developed a new protocol for embedding fluorescent beads in polymerization solution (Fig. 3b,e and Supplementary Note 11). Furthermore,

we combined ExM with IsoView light-sheet microscopy[4] allowing acquisition of multiview, multitile datasets of expanded tissues, which in turn enabled reconstruction of entire *Drosophila* larval nervous systems with a spatially isotropic subcellular resolution of ~100 nm (Fig. 3c,e). Image acquisition took 10 min as compared to the over 2 d that is required to image an expanded sample of similar volume using lattice light-sheet microscopy[24], although the acquisition was at lower resolution (Methods).

BigStitcher enables efficient and automatic processing of terabyte-sized datasets and addresses major unsolved issues such as easy import, management of large images, datasets acquired in a non-regular grid, globally optimal alignment of sparse datasets, illumination selection, rigid and non-rigid multiview alignment of multitile acquisitions, PSF extraction, quality estimation and interactive fusion. The aligned dataset and intermediate steps are interactively displayed, enabling the user to verify and interact with the alignment process to confirm and potentially guide proper alignment of complicated datasets (Supplementary Figs. 5, 16, 20 and 21). Automatic reconstruction of even large datasets can be achieved within tens of minutes, and BigStitcher outperforms existing software in terms of functionality, user-interaction and performance[12,13,16,24] (Supplementary Table 2; Methods). BigStitcher supports cleared samples (Fig. 3a,b,d), ExM samples (Fig. 3c,e and Supplementary Fig. 22), and standard 2D and 3D confocal and wide-field acquisitions, as well as tiled multiview light-sheet acquisitions (Fig. 3f). BigStitcher is open-source, implemented in ImgLib2 (ref. [10]) and provided as a Fiji[25] plugin with comprehensive documentation (https://imagej.net/BigStitcher). Most of its functionality is compatible with the ImageJ macro language (Methods) and can thus easily be automated. These properties make BigStitcher a powerful and scalable tool for the handling and reconstruction of tiled high-resolution image datasets acquired by new light microscopy technologies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0501-0.

Received: 18 June 2018; Accepted: 25 June 2019;
Published online: 05 August 2019

References

1. Chung, K. et al. *Nature* **497**, 332–337 (2013).
2. Chen, F., Tillberg, P. W. & Boyden, E. S. *Science* **347**, 543–548 (2015).
3. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E. H. K. *Science* **305**, 1007–1009 (2004).
4. Chhetri, R. K. et al. *Nat. Methods* **12**, 1171–1178 (2015).
5. Huisken, J. & Stainier, D. Y. R. *Opt. Lett.* **32**, 2608–2610 (2007).
6. Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. *Nat. Protoc.* **9**, 1682–1697 (2014).
7. Richardson, D. S. & Lichtman, J. W. *Cell* **162**, 246–257 (2015).
8. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. *Nat. Methods* **12**, 481–483 (2015).
9. Linkert, M. et al. *J. Cell Biol.* **189**, 777–782 (2010).
10. Pietzsch, T., Preibisch, S., Tomancák, P. & Saalfeld, S. *Bioinformatics* **28**, 3009–3011 (2012).
11. Kuglin, C. D. & Hines, D. C. in *Proc. IEEE International Conference on Cybernetics and Society* 163–165 (IEEE, 1975)
12. Preibisch, S., Saalfeld, S. & Tomancák, P. *Bioinformatics* **25**, 1463–1465 (2009).
13. Emmenlauer, M. et al. *J. Microsc.* **233**, 42–60 (2009).
14. Chalfoun, J. et al. *Sci. Rep.* **7**, 4988 (2017).
15. Saalfeld, S., Fetter, R., Cardona, A. & Tomancák, P. *Nat. Methods* **9**, 717–720 (2012).
16. Bria, A. & Ianello, G. *BMC Bioinformatics* **13**, 316 (2012).
17. Besl, P. J. & McKay, N. D. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (1992).
18. Migliori, B. et al. *BMC Biol.* **16**, 57 (2018).
19. Schaefer, S., McPhail, T. & Warren, J. *ACM Trans. Graph.* **25**, 533–540 (2006).
20. Ryan, D. P. et al. *Nat. Commun.* **8**, 612 (2017).
21. Royer, L. A. et al. Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nat. Biotechnol.* **34**, 1267–1278 (2016).
22. Nieuwenhuizen, R. P. J. et al. *Nat. Methods* **10**, 557–562 (2013).
23. Preibisch, S. et al. *Nat. Methods* **11**, 645–648 (2014).
24. Gao, R. et al. *Science* **363**, eaau8302 (2019).
25. Schindelin, J. et al. *Nat. Methods* **9**, 676–682 (2012).

Acknowledgements

We thank T. Pietzsch and S. Saalfeld for insightful discussions and BigDataViewer and ImgLib2 support; N. Vladimirov for very helpful microscopy discussions; C. Rueden for Fiji support and maintenance; N. Gompel for early-stage project discussions; and the *Caenorhabditis* Genetics Center at the University of Minnesota for providing *C. elegans* strains. S.P., F.P. and M.T. were funded by MDC Berlin; S.P. was supported by HFSP grant RGP0021/2018-102; F.P. was funded by a PhD fellowship from Studienstiftung des deutschen Volkes; F.R.R. and M.T. were funded by the Helmholtz Alliances ICEMED and AMPro; D.H., H.H. and H.L. were funded by the Deutsche Forschungsgemeinschaft (DFG, Nanosystems Initiative Munich), the NHGRI/NIH Center for Photogenomics (grant RM1 HG007743) and LMU Munich; and P.T., N.R., R.K.C., A.C. and P.J.K. were funded by HHMI Janelia.

Author contributions

S.P. conceived the idea in discussions with H.H., H.L. and M.T.; D.H. and S.P. developed the algorithms and implemented the software; F.R.R. performed all clearing experiments, reconstructions and benchmarks; F.P. imaged and reconstructed *C. elegans*; P.T. and N.R. performed ExM sample preparation; R.K.C. and P.J.K. developed the ExM-optimized IsoView microscope and imaged the sample; S.P. reconstructed the ExM sample; S.P., M.T., H.L., H.H., P.J.K. and A.C. supported and supervised the project; and S.P., D.H. and F.R.R. wrote the manuscript with input from the co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0501-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.P.

Peer review information: Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Animals. For clearing, we used a previously generated $Bsx^{H2BeGFP}$ mouse line[26], in which exon 1 of the *Bsx* gene is replaced (starting at the ATG) with a sequence encoding a histone 2B–enhanced green fluorescent protein (eGFP) fusion protein. Brains from 10-week-old female $Bsx^{H2BeGFP/+}$ mice were used for tissue clearing and imaging. *C. elegans* dauer larvae expressing the tagRFP fluorescent protein fused to a nuclear localization sequence in all neurons[27] (strain AML32) were obtained by selecting dauer larvae in 1% SDS for 30 min (ref. [28]). Dauer larvae were fixed with 4% paraformaldehyde for 30 min on ice, placed in 70% ethanol overnight at 4 °C and subsequently stained with DAPI. Experiments were conducted according to the institutional guidelines of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association after approval from the Berlin State Office for Health and Social Affairs (LAGeSo, Landesamt für Gesundheit und Soziales). *Drosophila* larvae used for ExM were obtained from the strain w;;attp2, carrying an empty attp2 landing site[29].

Clearing and expansion. Tissue clearing was performed using the CLARITY protocol[1] (Supplementary Note 1). Before imaging, the tissue samples were placed overnight in FocusClear for refractive index matching. For ExM, the nervous system of a first-instar *Drosophila* larva was extracted, fixed and stained with anti-tubulin antibodies (Supplementary Note 2). The stained sample was washed in 1× PBS and then processed using a modified ExM method to achieve 7.5-fold expansion in each dimension (Supplementary Note 2). In summary, the specimen was treated with acryloyl-X[30] as in standard ExM and embedded using a gel recipe modified from the original method[2]. The modified recipe uses a reduced cross-linker concentration to achieve greater expansion. After digestion with proteinase K, a new re-embedding step toughens up the gel, which would otherwise have poor mechanical properties.

Imaging. 3D images of cleared mouse brains placed in a FocusClear-filled imaging chamber were acquired using the Zeiss light-sheet Z.1 microscope. Fixed *C. elegans* dauer larvae were embedded in agarose that contains fluorescent beads, and imaged using the same microscope in a water-filled sample chamber. 3D images of the cleared and expanded central nervous system of a *Drosophila* first-instar larva were acquired using an IsoView light-sheet microscope[4] that has been modified for multitile acquisition. The ability of the IsoView microscope to rapidly record large specimens allowed us to image the entire 7.5-fold expanded nervous system ($2{,}040 \times 1{,}108 \times 1{,}201\ \mu m = 2.7\ mm^3$) in 10 min with an isotropic resolution of approximately 100 nm, acquiring two volumes that were oriented orthogonally to each other each with a spatial sampling of $55 \times 55 \times 110$ nm (unexpanded volume). In comparison, acquisition of a 4.09-fold expanded volume of half the size ($1{,}400 \times 2{,}700 \times 370\ \mu m = 1.4\ mm^3$) with two channels using lattice light-sheet microscopy[24] took 2.61 d; however, the image acquired using lattice light-sheet microscopy had a higher spatial sampling of $23.7 \times 23.7 \times 44.0$ nm (unexpanded volume). Details of the imaging strategies are described in Supplementary Note 3 and a summary of the most important acquisition parameters can be found in Supplementary Table 1.

Data processing pipeline. All data shown in this paper were processed using the BigStitcher Fiji[25] plugin. Zeiss CZI files and TIFF files exported by custom microscopes were imported using AutoLoader and subsequently converted to HDF5. For Zeiss CZI files, approximate tile positions and rotation angles were imported automatically; for other files, these parameters were specified by hand using BigStitcher tools (Supplementary Figs. 5 and 21). If not stated otherwise, reconstruction was performed using the following steps. For each tile, the best illumination was selected. Tiles were aligned using the phase correlation method using two-round global optimization followed by ICP[17] refinement on an affine model. Interest-point detection was performed for each multitile view. Either our extension of the fast descriptor-based rotation-invariant algorithm[31] or the descriptor-based translation-invariant algorithm[32], which was used after application of manual rotation, was used to register the interest points of each angle, and another round of ICP refinement was performed on all image tiles of the acquisition. Fused and deconvolved images were exported as TIFF files.

Pairwise stitching using Fourier-based phase correlation. We calculated pairwise translational shifts using our ImgLib2 (ref. [10]) implementation of the Fourier-based phase correlation algorithm[11]. The processing time was substantially reduced, while our simulations showed that, at the same time, registration errors below 1 pixel were achieved by computing the phase correlation on downsampled images and performing subpixel localization of the shift vector[33] (Supplementary Notes 6 and 7, and Supplementary Figs. 11–15).

Global optimization. To calculate the final transformations of each image tile we extended the concept of globally optimal registration by iterative minimization of square displacement of point correspondences[12,15,31,32] (Supplementary Note 12). Erroneous pairwise links that might not have been filtered out before global optimization (for example, wrong links caused by average cross-correlation, repetitive patterns or a low number of correspondences in the ICP refinement) lead to high registration errors after global optimization. This manifests in a large distance error, which is the difference between the individually computed distance between images (link) and the actual distance between images after global optimization. Iterative removal of the link with the highest distance error from the link graph and repetition of the global optimization lead to convergence to user-defined thresholds[12]. We extended this concept from Preibisch et al.[12] to affine transformations, introduced a new heuristic that additionally incorporates link quality, and implemented it in an extendable framework required for the two-round global optimization (Supplementary Note 12). If the dataset contains empty tiles or even consists of multiple disconnected objects, the final transformations are not propagated between them (Supplementary Fig. 10). We therefore developed a two-round global optimization that is capable of aligning independent connected components of the link graph using weak links defined by the current transformations of each tile (for example, approximate locations from metadata or manual alignments), which optimally preserves distances between neighboring objects (Supplementary Note 12 and Supplementary Fig. 10).

Our global optimization is agnostic to the nature of the point correspondences and transformation model, which allows us to use the same algorithm for translation-based alignment of tiled datasets using phase correlation, as well as affine registrations of multitile, multiview datasets based on ICP refinement or geometric descriptor matching.

Iterative closest point refinement. Although the phase-correlation-based image stitching produces relatively high-quality alignments, smaller errors can remain (Supplementary Figs. 13–15). Furthermore, this method is not able to correct for non-translational effects such as chromatic and spherical aberration or sample-induced light refraction (Fig. 2 and Supplementary Figs. 3, 4, 9 and 17). These effects can be better approximated using affine transformations. We therefore automatically detect interest points and run an ICP algorithm[17] for each overlapping pair of images, where the assignment of correspondences is limited by a distance threshold. We used the identified corresponding points of all pairwise links and computed a globally optimal affine transformation for each tile using our new global optimization algorithm. To avoid scaling of datasets, we regularized the affine transformation using a rigid transformation[15]. The resulting alignment usually improved the alignment quality and the same strategy could be applied to multichannel alignment if the same autofluorescent structures were visible in multiple channels (Supplementary Fig. 3). However, only small chromatic aberrations within a few pixels could be corrected by approximations based on affine or split-affine transformations (Supplementary Fig. 3). At the same time, non-rigid transformations can easily be unstable for this purpose as correspondences between different channels would have to be distributed over the entire image. Therefore, in those cases, it is better to use specialized software such as that described by Matsuda et al.[34] to correct for chromatic aberrations before importing data into BigStitcher.

Illumination selection. When imaging large samples using sequential dual-sided illumination[5], typically only illumination from one direction provided good image quality (Supplementary Fig. 2). We therefore implemented an illumination selection functionality in BigStitcher. We first combined all (selected) images by their illumination attribute and thereby grouped all images that shared other attributes besides illumination direction. In each of the resulting groups, we selected the best image. As quality criteria, we offered either the rFRC on full-resolution images or a fast approximation obtained using mean intensity or mean gradient magnitude at the lowest resolution level. While rFRC provided the highest distinctive power, both fast approximations were typically sufficient for robust estimation of the higher-quality illumination direction (Supplementary Fig. 2 and Fig. 1d). The image with the highest quality score was kept, while all other images were marked as missing in the dataset, which led to them being ignored in subsequent processing steps. However, before applying automatic illumination estimation, the user has the option to verify and potentially change the result. Optional resaving of the dataset after this step potentially decreases the storage requirement twofold.

Simulation of light propagation in tissue using ray tracing. We observed non-rigid deformations occurring in areas where image data from opposing light sheets were recorded by the same camera (Fig. 2a). While it is intuitive and clear from existing simulations of light propagation[35] that imaging of the same light sheet from opposing objectives can lead to non-rigid deformations caused by different aberrations in the detection light paths, we wanted to understand how changing just the light-sheet direction can introduce non-rigid deformations. These effects were clearly visible in large samples like entire mouse brains (Fig. 2a), which are still beyond the range of simulation using reasonable efforts. We therefore developed a simple ray-tracing-based method to simulate light propagation, which aimed to recapitulate these observations at a smaller and more manageable scale (Supplementary Note 13).

Although quite simple, this simulation recapitulated the effects observed in cleared images (Fig. 2a,h) and illustrated that different refraction of the illumination light sheets alone can lead to non-rigid deformations in the acquired image stacks as it lead to illumination of different contents of the sample in the *z* axis (Fig. 2g,h and Supplementary Video 2). As detection objectives with relatively

low numerical aperture (NA; and therefore large depth of field) are typically used for detection (Supplementary Table 1), both light sheets can still appear in focus despite being tens of micrometers away from each other. Therefore, these transformations need to be corrected for.

Non-rigid transformation. To be able to compensate for potentially strong aberrations introduced by light refraction due to the sample, we implemented a non-rigid alignment method. It is based on the concept of moving least squares[19] that has been shown to perform well in biomedical applications[15]. We implicitly regularized moving least squares using ICP[17] or random sample consensus[36] (Supplementary Note 14). Moving least squares requires computation of a transformation for each pixel, which is computationally expensive. We therefore implemented a virtual cached layer that only computes a transformation for every m^{th} pixel (m being the distance between pixels for which transfomations are computed) and linearly interpolates affine transformations for pixels in between. As BigDataViewer currently only supports affine transformations, we additionally implemented a multiresolution preview based on virtually fused non-rigid volumes that can be interactively displayed in an extra BigDataViewer window. BigStitcher also supports 'hybrid' fusion of non-rigid- and affine-transformed image tiles as non-rigid registration requires substantially increased computational effort (Supplementary Fig. 17).

Virtual reblocking. To allow piecewise affine transformations or a more refined illumination selection, we developed virtual reblocking of all 3D image stacks of an acquisition (Fig. 2l,m). The implementation distributes 3D blocks onto each input image stack using a defined overlap and thereby defines a new set of input image stacks for the acquisition. The new virtual image stacks are computed on the fly for all resolution levels from the original image data. Any ImgLoader is supported, including multiresolution image stacks. Optional resaving of the dataset as HDF5 or TIFF transforms the virtual image stacks into physical representations.

Quality estimation on the basis of Fourier ring correlation. To estimate image quality across entire terabyte-sized acquisitions, we developed an extension of the FRC[22] that is robust and insensitive to camera noise. $\mathrm{FRC}_{r_1,r_2}(f)$ constitutes the per-spatial-frequency (f) correlation between two independent realizations, r_1 and r_2, of the same signal. In localization-based super-resolution, point clouds are therefore typically split into two sets of independent pixels. Here we use consecutive slices instead and take advantage of the fact that they are nearly identical owing to the axial extent of the PSF. As a result, $\mathrm{FRC}_{r_1,r_2}(f)$ constitutes an entire correlation spectrum for each z plane, and we compute a single quality value $Q(z)$ by integration over all frequencies f

$$Q_{\mathrm{FRC}}(z) = \int_f \mathrm{FRC}_{z,z+1}(f)$$

A smoother symmetric result can be obtained by averaging the FRC spectra obtained using z planes above and below the measured plane

$$Q_{\mathrm{FRC}}(z) = \int_f \frac{\mathrm{FRC}_{z,z+1}(f) + \mathrm{FRC}_{z,z-1}(f)}{2}$$

For computing the 2D FRC, we adapted methods from the FRC ImageJ plugin (Supplementary Note 15). This results in a precise estimation of image quality, except if patterned noise (for example, sCMOS camera noise) is the dominant source of signal (Supplementary Fig. 1). To overcome this instability, we developed the rFRC, which subtracts a locally estimated scatterplot smoothing (loess)-smoothed[37] baseline FRC of z planes spaced by m slices that are beyond the axial extent of the PSF

$$Q_{\mathrm{rFRC}}(z) = \int_f \frac{\mathrm{FRC}_{z,z+1}(f) + \mathrm{FRC}_{z,z-1}(f)}{2} - \mathrm{loess}\left(\mathrm{FRC}_{z-m,z+m}(f)\right)$$

This effectively measures which additional frequencies the central planes z, $z+1$ and $z-1$, z have in common, as compared to the planes $z-m$, $z+m$ that are beyond the axial size of the PSF. The resulting values $Q_{\mathrm{rFRC}}(z)$ robustly measure image quality in the sample (Supplementary Figs. 1g, 7 and 8). As image content can change drastically within a slice, we support computation using a manually defined block-size (for example, 512×512) and with z stepping (for example, every 20 planes). To estimate the quality metric for entire acquisitions, we compute $Q_{\mathrm{rFRC}}(z)$ for defined points in each image stack. Over all input stacks, these measurements are held as sparse representations using ImgLib2 (ref. [10]) that can be rendered virtually and overlaid onto entire fused volumes (Supplementary Figs. 7 and 8, and Supplementary Video 8 and 9).

Virtual image fusion. A set of overlapping transformed image tiles are fused into one output image using a per-pixel weighted average that minimizes boundary artifacts and can increase contrast by incorporating entropy estimation[38] (Supplementary Note 16). To correct for unequal brightness and contrast in adjacent images, we optionally perform adjustment of the pixel intensities using a linear transformation for each image. An optimal brightness and contrast adjustment can be estimated using the same optimization framework used for image registration[39] (Supplementary Note 17). The memory requirements for fusion of large volumes can easily exceed the available random access memory (RAM) on a machine owing to the size of the output and the combined size of the input images. We, therefore, developed a framework for intensity transformations and coordinate transformations that is based on ImgLib2 (ref. [10]) and virtually fused all pixels of a defined bounding box using all input images and their associated weights. As the input images are provided through virtual image loading, the size of a virtually fused image is close to zero, irrespective of the size of input and output images. Ideally, input images are available in blocks so that affine transformations that slice the image stacks in arbitrary orientations do not require loading of the entire image[8]. The output image can now be rendered on a pixel-by-pixel basis with minimal memory requirements. Additional caching of the input image and the output images allows an efficient multithreaded fusion for optimally fast processing given the available memory. Therefore, more RAM will effectively speed up the fusion process (Supplementary Table 2 and Supplementary Fig. 17), but even machines with very low RAM are able to fuse terabyte-sized volumes (Supplementary Fig. 19). Fused images can be saved by choosing cached or virtual fusion and subsequently saving the ImageJ virtual stack using 'Save as image sequence…'.

Downsampling of the output can easily be incorporated by scaling the bounding box and preconcatenation of the downsampling transformation with each image transformation. If the input files are multiresolution, we automatically compute the optimal resolution level at which the input needs to be loaded. To optionally further reduce the image size of the fused image, the graphical user interface offers to conserve the original anisotropy between lateral and axial orientations of the acquired sample, which is a sensible choice if the dataset contains a single view or opposing (for example, 0° and 180°) multitile views.

Macro automation and headless operation. In addition to the graphical user interface, we offer standalone Fiji plugins for most of the individual steps, such as data import, illumination selection, pairwise shift calculation, link filtering, multiview alignment, global optimization, and image fusion and deconvolution. In macro mode results will not be displayed interactively but are instead saved to the XML project file or output files immediately. The individual steps can be recorded as ImageJ macros and easily combined into a script for headless batch processing[40].

Limitations of the framework and other software solutions. BigStitcher is designed for the reconstruction of large multitile, dual-illumination, multiview datasets, and supports affine, split-affine and non-rigid registrations to solve the alignment process for terabyte-sized image data. Several solutions based on image correlation support multitile-only data, such as Terastitcher[16], XUVTools[13] and ImageJ Stitching[11]; however, these solutions only support translation models, making them unsuitable for the types of datasets described above (Fig. 2 and Supplementary Figs. 3, 4, 9 and 17). A recent stitching solution developed by the Saalfeld lab that is also based on ImgLib2 (ref. [10]) can handle even larger datasets (up to hundreds of terabytes) and supports affine transformations based on local cross-correlation[24]. It is, however, also focused on multitile-only acquisitions, is designed to run on a cluster or in the cloud, does not support non-linear tile deformations, and has no user interface to access its functionality. Currently, BigStitcher scales well up to 1,000 large 3D image tiles per time point and image sizes on the order of tens of terabytes per time point. This is, however, mostly due to a limit in the rendering capacity of BigDataViewer. Future optimizations of BigDataViewer[8] and/or BigStitcher will further increase this limit. BigStitcher can correct for chromatic and spherical aberrations by approximation with affine transformations if errors are within a few pixels. For chromatic aberration correction, enough autofluorescent structures must be visible across channels (Supplementary Figs. 3, 4 and 9). Although BigStitcher can correct for geometric transformations introduced by the acquisition process (Fig. 2 and Supplementary Fig. 17), it is not possible to correct for images that are out of focus.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Small example datasets are available for download from the Open Science Foundation at https://osf.io/bufza/. Larger datasets are available on request. Additional datasets uploaded at a later stage will be linked from the documentation page which can be found at https://imagej.net/BigStitcher#Example_Datasets. Example datasets are explained in detail in Supplementary Note 18.

Code availability

All source code used in this publication (BigStitcher, phase correlation simulation and benchmarks, and the simulation of light propagation in tissue using ray tracing) is open-source and published under the GNU General Public License version 2. The latest stable releases used in this publication are provided as

Supplementary Software; current versions that include bugfixes and updates can be downloaded from GitHub (at https://github.com/PreibischLab/BigStitcher; https://github.com/PreibischLab/multiview-reconstruction; and https://github.com/PreibischLab/multiview-simulation; see Supplementary Notes 19 and 20 for further explanations). Details on how to use the software are described in Supplementary Note 21.

References

26. Sakkou, M. et al. *Cell Metab.* **5**, 450–463 (2007).
27. Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W. & Leifer, A. M. *PLoS Comput. Biol.* **13**, e1005517 (2017).
28. Karp, X. in *WormBook* (ed. The *C. elegans* Research Community, 2016); https://doi.org/10.1895/wormbook.1.180.1
29. Pfeiffer, D. B. et al. *Proc. Natl Acad. Sci. USA* **105**, 9715–9720 (2008).
30. Tillberg, P. W. et al. *Nat. Biotechnol.* **34**, 987–992 (2016).
31. Preibisch, S., Saalfeld, S., Schindelin, J. & Tomancák, P. *Nat. Methods* **7**, 418–419 (2010).
32. Smith, C. S. et al. *J. Cell Biol.* **209**, 609–619 (2015).
33. Lowe, D. G. *Int. J. Comput. Vis.* **60**, 91–110 (2004).
34. Matsuda, A., Schermelleh, L., Hirano, Y., Haraguchi, T. & Hiraoka, Y. *Sci. Rep.* **8**, 7583 (2018).
35. Weigert, M., Subramanian, K., Bundschuh, S. T., Myers, E. W. & Kreysing, M. *PLoS Comput. Biol.* **14**, e1006079 (2018).
36. Fischler, M. A. & Bolles, R. C. *Commun. ACM* **24**, 381–395 (1981).
37. Cleveland, W. S. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
38. Preibisch, S., Rohlfing, T., Hasak, M. P. & Tomancák, P. in *Proc. of the International Society for Optics and Photonics, Medical Imaging* (eds. Reinhardt, J. M. & Pluim, J. P. W.) (SPIE, 2008).
39. Blasse, C. et al. *Bioinformatics* **33**, 2563–2569 (2017).
40. Schmied, C., Steinbach, P., Pietzsch, T., Preibisch, S. & Tomancák, P. *Bioinformatics* **32**, 1112–1114 (2016).

nature research

Corresponding author(s): Stephan Preibisch

Last updated by author(s): May 28, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Simulated data was created using the net.preibisch.simulation.SimulateMultiViewAberrations class in the multiview-simulation package (release version 0.2.2). Since it is a Maven artifact, the versions of all dependencies are clearly defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/multiview-simulation/commit/b41b74cce9287f804b670d7de3396605446818a8).
Data analysis	The data was reconstructed using BigStitcher (release version 0.3.3). Since it is a Maven artifact, the versions of all dependencies (e.g. multiview-reconstruction) are clearly defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/BigStitcher/commit/0d7f79a59ab15fb1805157ab72c5bc9802b02fbd).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Small example datasets are available online: https://osf.io/bufza/ (Open Science Foundation). Larger datasets are available on request. When we find ways to host larger datasets, they will be linked from the documentation page: https://imagej.net/BigStitcher#Example_Datasets

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size of 8 acquired samples was chosen to show a wide variety of samples and sample preparation that can be reconstructed using BigStitcher. Some samples are quite similar (clearing), some were chosen to highlight the generality (expanded, only fixed).
Data exclusions	No data was excluded.
Replication	In total 8 different samples (cleared, expanded, only fixed) were acquired and reconstructed for this publication. Furthermore, one sample was simulated and reconstructed. Additionally, the BigStitcher reconstruction pipeline has been applied to >50 samples in our lab (not part of the publication, most are prepared in collaboration on other projects), which highlights the generality of the method. The tissue sections shown in Fig. 3a + Suppl. Fig. 1 + Suppl. VIdeo 1 as well as Fig. 3b are very similar and highlight reproducibility.
Randomization	Randomization was used for a) the simulation and benchmarking of phase correlation (Supplementary Fig. 12-15), as well as b) for the raytracing (Fig. 2b-e). For a) random overlaps (uniformly distributed) were simulated 300 times for each condition, and the noise for the image generation process is based on Poisson noise. For b), illumination and detection rays in defined areas are sent randomly (uniformly distributed) into the scene.
Blinding	Blinding is not applicable for this study since there is no process that requires it.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For clearing we used a previously generated BsxH2BeGFP mouse line, where the exon 1 of the bsx gene is replaced starting at the ATG with the coding sequence for histone2B eGFP. Brains from 10-week old female BsxH2BeGFP/+ mice were used for tissue clearing and imaging. C. elegans dauer larvae expressing tagRFP fused to a nuclear localizing sequence under the pan-neuronal rab-3 promotor in all neuron nuclei were obtained by selecting dauer larvae in 1% SDS for 30 minutes. Drosophila larva used for ExM were obtained from the strain w;;attp2, carrying an empty attp2 landing site.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	Experiments were conducted according to institutional guidelines of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association after approval from the Berlin State Office for Health and Social Affairs (LAGeSo, Landesamt für Gesundheit und Soziales, Berlin, Germany)

Note that full information on the approval of the study protocol must also be provided in the manuscript.