

# An integrated platform for high-throughput nanoscopy

Received: 2 June 2022

Accepted: 2 February 2023

Published online: 13 March 2023



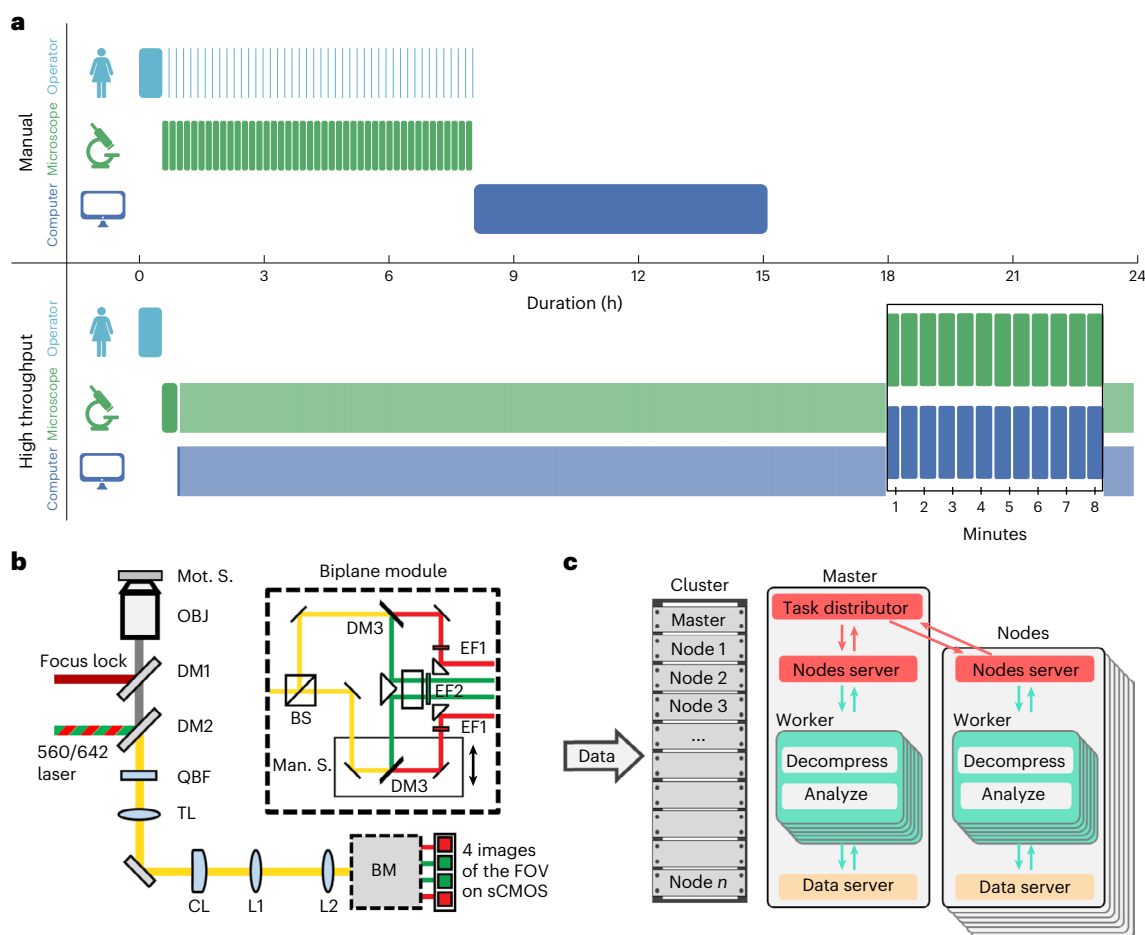
Andrew E. S. Barentine<sup>1,2,9</sup>, Yu Lin<sup>1,2,9</sup>, Edward M. Courvan<sup>1,3</sup>, Phylcia Kidd<sup>1</sup>, Miao Liu<sup>4</sup>, Leonhard Baldof<sup>1,5</sup>, Timy Phan<sup>1,5</sup>, Felix Rivera-Molina<sup>1</sup>, Michael R. Grace<sup>1</sup>, Zach Marin<sup>1,2,6</sup>, Mark Lessard<sup>1</sup>, Juliana Rios Chen<sup>1</sup>, Siyuan Wang<sup>1,4</sup>, Karla M. Neugebauer<sup>1,3</sup>, Joerg Bewersdorf<sup>1,2,7,8</sup>✉ & David Baddeley<sup>1,6,8</sup>✉

Single-molecule localization microscopy enables three-dimensional fluorescence imaging at tens-of-nanometer resolution, but requires many camera frames to reconstruct a super-resolved image. This limits the typical throughput to tens of cells per day. While frame rates can now be increased by over an order of magnitude, the large data volumes become limiting in existing workflows. Here we present an integrated acquisition and analysis platform leveraging microscopy-specific data compression, distributed storage and distributed analysis to enable an acquisition and analysis throughput of 10,000 cells per day. The platform facilitates graphically reconfigurable analyses to be automatically initiated from the microscope during acquisition and remotely executed, and can even feed back and queue new acquisition tasks on the microscope. We demonstrate the utility of this framework by imaging hundreds of cells per well in multi-well sample formats. Our platform, implemented within the PYthon-Microscopy Environment (PYME), is easily configurable to control custom microscopes, and includes a plugin framework for user-defined extensions.

Super-resolution single-molecule localization microscopy (SMLM) offers a roughly tenfold improvement in resolution over conventional, diffraction-limited fluorescence microscopy, but it does so at the expense of acquisition time, data volume and analysis overhead<sup>1</sup>. For SMLM techniques such as (d)STORM/(F)PALM (stochastic optical reconstruction microscopy/photoactivated localization microscopy), a single region of interest (ROI) a few tens of micrometers in diameter usually requires a series of 10,000 to 100,000 camera raw data frames. These are typically acquired at 50 frames per second (FPS), meaning that a day of diligent manual imaging yields only a few tens of fields of view (FOVs) (Fig. 1a).

As a result, the majority of SMLM applications giving rise to new biological insight have addressed questions that can be answered with just a handful of super-resolved images, either because the structure is obvious from few observations, or because a large number of structures are observable in the FOV<sup>2–5</sup>. A notable, and impressive, exception to typical sample sizes is the study by Boettinger et al. on gene compaction<sup>6</sup>, which manually imaged over 2,000 eukaryotic cells and would have required several months of data collection. Efforts to automate SMLM image acquisition<sup>7–9</sup> have reduced the amount of time the operator needs to spend in front of the microscope, but have not been combined with large improvements in acquisition speed.

<sup>1</sup>Department of Cell Biology, Yale School of Medicine, New Haven, CT, USA. <sup>2</sup>Department of Biomedical Engineering, Yale University, New Haven, CT, USA. <sup>3</sup>Department of Molecular Biophysics and Biochemistry, Yale School of Medicine, New Haven, CT, USA. <sup>4</sup>Department of Genetics, Yale School of Medicine, New Haven, CT, USA. <sup>5</sup>Department of Computer Science and Mathematics, University of Applied Sciences, Munich, Germany. <sup>6</sup>Auckland Bioengineering Institute at University of Auckland, Auckland, New Zealand. <sup>7</sup>Department of Physics, Yale University, New Haven, CT, USA. <sup>8</sup>Nanobiology Institute, Yale University, West Haven, CT, USA. <sup>9</sup>These authors contributed equally: Andrew E. S. Barentine, Yu Lin. ✉e-mail: [joerg.bewersdorf@yale.edu](mailto:joerg.bewersdorf@yale.edu); [d.baddeley@auckland.ac.nz](mailto:d.baddeley@auckland.ac.nz)



**Fig. 1 | High-throughput SMLM. a**, Example timelines for SMLM acquisition of 36,000-frame ROIs performed at 50 FPS manually and automated at 800 FPS. **b**, Schematic of automated multicolor 3D biplanar-astigmatism SMLM microscope. Mot. S., motorized sample stage; OBJ, objective; DM1–3, dichroic

mirrors; QBF, quad-band filter; TL, tube lens; CL, cylindrical lens; L1–2, relay lenses; BM, biplane module; BS, beamsplitter cube; Man. S., manual translation stage; EF1–2, emission filters. **c**, Diagram of scalable data pipeline for real-time localization and automated post-localization analysis.

By using sCMOS (scientific complementary metal oxide semiconductor) cameras<sup>10</sup> and high laser intensities it is possible to acquire SMLM data an order of magnitude faster<sup>11,12</sup> than the typical 50 FPS, without a major loss in data quality. However, automating SMLM imaging at these high frame rates generates data at a rate of 800 MB s<sup>-1</sup>, posing unique challenges for both data storage and analysis. The analysis burden is further compounded by the need to account for sCMOS-specific noise characteristics (which are more complex than those of EMCCDs (electron multiplying charge-coupled devices)<sup>10</sup>) to obtain high-quality localization data. As a result, fast (>400 FPS) SMLM imaging has been largely restricted to two dimensions and has entailed a number of manual steps, in both image acquisition and analysis<sup>10,11,13</sup>. To deliver a truly high-throughput automated SMLM platform operating at sCMOS speeds represents a technological challenge requiring advances across microscope hardware, data handling and analysis routines.

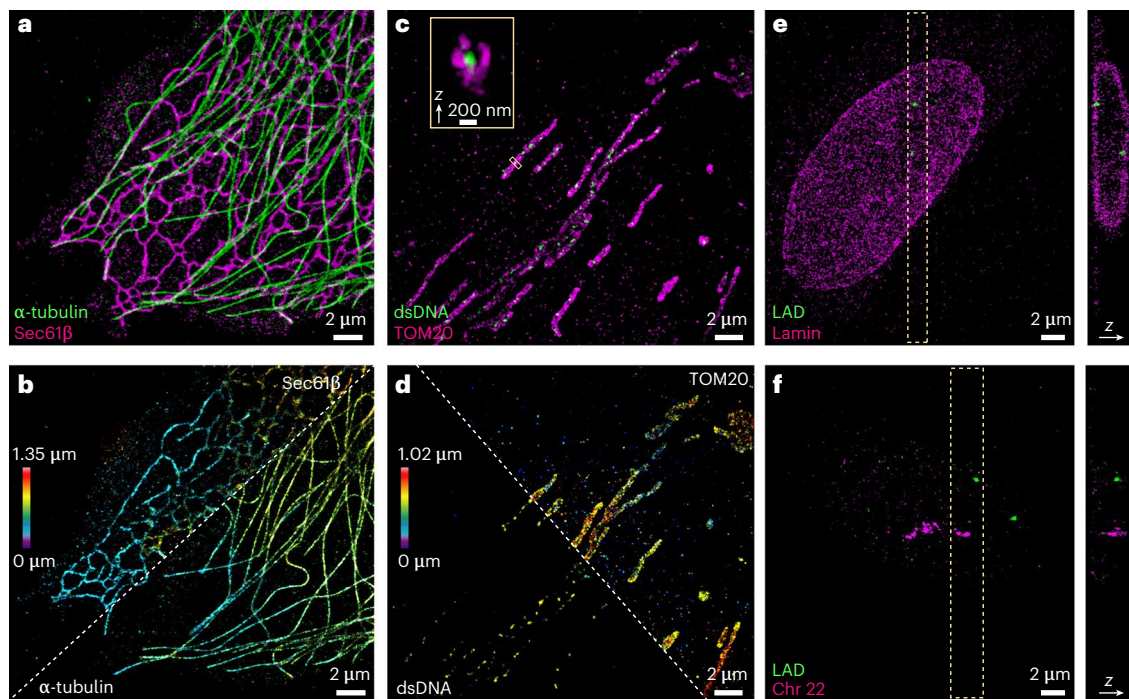
Here we present an integrated high-throughput SMLM platform operating at sCMOS speeds, transforming SMLM from an imaging technique specialized for small sample sizes into a high-throughput quantitative tool. We leverage our developments in data compression, distributed storage and distributed analysis to automatically perform real-time localization analysis, and additionally present a flexible architecture for distributed and automatic post-localization analysis and feedback-based imaging workflows. Our multicolor three-dimensional (3D) SMLM system is capable of imaging 10,000 mammalian cells a day, or entire studies configured on multi-well plates. We achieved a volumetric throughput of approximately 800,000  $\mu\text{m}^3 \text{h}^{-1}$  in two colors

compared with approximately 4,000  $\mu\text{m}^3 \text{h}^{-1}$  (single-color) for earlier automated 3D SMLM approaches<sup>8</sup>, based on 3D ROI sizes, number of ROIs imaged and total acquisition time. The microscope control, data storage and analysis pipeline integration is readily accessible to the community through the open-source PYthon-Microscopy Environment (PYME) (<http://python-microscopy.org>), which additionally features advanced visualization<sup>14</sup> and plugin extensibility, making it a viable tool for complete and customized SMLM workflows.

## Results

### Imaging system

We built a microscope with hardware optimized for automated high-speed single-molecule imaging. It features an sCMOS camera capable of capturing a 2,048 × 256-pixel ROI at 800 Hz, high-power lasers, motorized lateral and axial stages, and a focus stabilization system. Custom spectral and focal splitting optics (Fig. 1b, Supplementary Fig. 1 and Supplementary Note 1.1) allow us to image two spectral channels, each at two different focal planes, simultaneously. By increasing the offset between focal planes from a typical biplane configuration<sup>15,16</sup> to 750 nm and adding astigmatism<sup>17</sup> we can achieve high-quality 3D localization of single molecules over an extended axial range of about 1.2  $\mu\text{m}$ . This halves the number of axial steps which need to be taken when performing 3D volumetric imaging, resulting in a corresponding improvement in speed. Using this setup, we can acquire multicolor 3D super-resolved images in about 10 s, as shown with several examples in Fig. 2 (see also Supplementary Fig. 2).



**Fig. 2 | 3D multicolor acquisition at 800-Hz frame rate.** **a**, (colored by label), **b**, (colored by z position), Rapid two-color 3D SMLM of microtubules ( $\alpha$ -tubulin immunolabeled with CF568) and endoplasmic reticulum (Sec61 $\beta$ -GFP immunolabeled with AF647) in a COS-7 cell. **c**, (colored by label), **d**, (colored by z position), Mitochondria (TOM20 immunolabeled with AF647) and nucleoids (dsDNA immunolabeled with CF568ST) in a U-2 OS cell. **e**, Lamin A/C (immunolabeled

with CF568) and a LAD (Chr13:24405079-24709084, labeled with AF647 via FISH) in an IMR-90 cell. **f**, All 27 TADs along chromosome 22 (FISH-labeled with CF568) and a LAD (Chr5:115508197-115813276, FISH-labeled with AF647) in an IMR-90 cell. All datasets were acquired at a frame rate of 800 Hz for 8,000 frames (**a–d**) or 24,000 frames (**e,f**). FISH, fluorescence in situ hybridization.

### Scalable data handling—compression, streaming and distributed storage

Imaging at the full frame rate of an sCMOS camera generates an enormous amount of data ( $800 \text{ MB s}^{-1}$ ) and sustained imaging at this data rate is nontrivial. Previous high-speed SMLM efforts have saved data directly to a local solid-state drive (SSD)<sup>10</sup>, but this limits acquisitions to a few hours before the SSD is full—even on the largest SSDs currently available. Once full, copying the data to slower storage (for example, hard disk drives (HDDs)) can take substantially longer than the acquisition itself. To address the data movement and storage bottlenecks imposed by our high data rate, we developed a compression algorithm optimized for our microscope's noise model, and distributed storage across a small computer cluster (Figs. 1c and 3a and Supplementary Note 2).

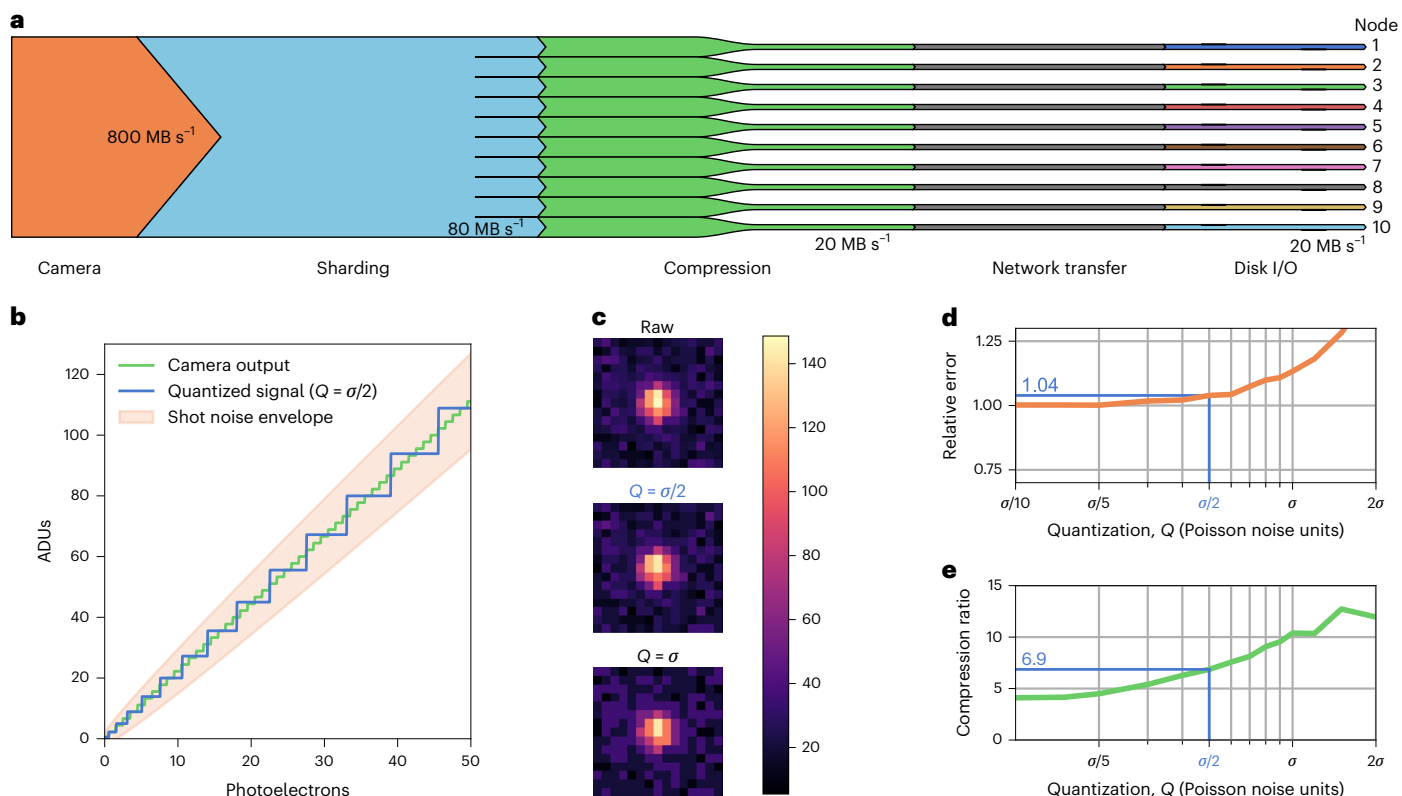
Standard lossless compression algorithms such as zip offer a modest 2–3-fold reduction in file size when applied to SMLM raw data and are typically not fast enough to allow real-time compression at  $800 \text{ MB s}^{-1}$ . These algorithms use entropy coding (for example, Huffman coding) which looks at the histogram of the data and uses short codes for frequently encountered values and longer codes for less frequently encountered values, combined with algorithms which encode repeated patterns. The poor compression ratios can be explained by two factors: SMLM raw data contain little repetitive structure for compression algorithms to exploit, in part due to Poisson noise, and the data as they come from the camera are very conservatively quantized with more unique values than necessary to accurately represent the data; data from sCMOS cameras are typically quantized such that one photoelectron corresponds to  $\sim 2.5$  analog-to-digital units (ADUs). This is reasonable for low signals (1–2 photoelectrons) where Poisson noise (which scales as  $\sigma(N) = \sqrt{N}$ ) is approximately  $\pm 1$  photoelectron (or approximately  $\pm 2.5$  ADUs). However, a signal of 100 photoelectrons will have an error of  $\pm 10$  photoelectrons (approximately

$\pm 25$  ADUs), giving rise to a band of 50 unique values which are not meaningfully distinct but will nonetheless make the signal harder to compress. The solution is simple: rather than using a constant quantization interval as provided by the camera, we re-quantize our data before compression such that the interval between quantal units is a constant fraction,  $Q$ , of the expected Poisson noise rather than a constant number of ADUs (depicted in Fig. 3b).

We systematically varied  $Q$  for simulated SMLM raw data (Fig. 3c and Supplementary Note 3) and analyzed the resulting relative localization error for obtained localizations (Fig. 3d) and the achieved data compression ratio (Fig. 3e). At  $Q = \sigma/2$ , we achieve 6.9-fold compression at a relative localization error of only 4%. This  $>5$ -fold compression ratio holds across a large range of emitter densities (Supplementary Fig. 3).

While this re-quantization is technically lossy, it ensures that any losses are within the original data noise envelope (Fig. 3b) and also visually preserves the integrity of blinking emitters, as seen in Fig. 3c. Combined with the low relative localization error, we therefore deemed this to be an acceptable compromise and used  $Q = \sigma/2$  in our further experiments.

Noting that the repeated pattern encoding portion of compression algorithms is not very helpful for our data, we skip this step entirely, greatly improving our speed (Supplementary Fig. 4). The resulting compression is performed in real-time on the microscope computer and spooled over the network to a small computer cluster where each node is running PYME server processes (Fig. 1c and Supplementary Fig. 6). By sharding the data across multiple nodes on a per-frame basis, we further decrease the amount of per-disk bandwidth, and can use HDDs instead of SSDs for a low-cost, high-volume storage solution. This combination of noise model-aware compression and distributed storage thereby enables us to stream continuously at the full camera frame rate and also lays the foundation for analyzing the data in parallel.



**Fig. 3 | Data volume solutions.** **a**, Sankey diagram showing approximate data bandwidths as they are transferred from the camera to instrument computer RAM before being sharded, compressed (lossy), sent across a local network and saved locally on HDDs on multiple computer nodes. **b**, Our lossy compression algorithm re-scales the ADUs such that the corresponding number of

photoelectrons represented by each unique value scales with a set fraction of the shot noise. **c**, A simulated localization ROI shown at various quantization levels. **d**, The relative localization error as a function of quantization for simulated localizations with an sCMOS noise model. **e**, The compression ratio achieved at the same quantization levels as in **d**.

## Real-time localization

Waiting for hours or days to do quality control or analysis on localization results largely defeats the purpose of imaging at high bandwidths. Although some simplified methods (such as centroiding) have been shown to yield high localization speed, these entail serious compromises in precision and accuracy<sup>18</sup>. Obtaining optimal localization precision with sCMOS cameras requires fitting using an sCMOS-specific noise model<sup>10</sup> and an algorithm such as weighted least squares<sup>19</sup> or maximum likelihood estimation<sup>20</sup>.

We optimized our previous graphical processing unit (GPU)-based maximum likelihood estimation code<sup>10</sup> to achieve about 15-fold improvement in speed by using one thread per localization ROI pixel to evaluate the model function, with a single thread performing parameter updates on each iteration, rather than a single thread doing the entire fit (Supplementary Fig. 7). We also accelerated sliding-window background estimation and candidate molecule detection by moving them from the CPU to the GPU, parallelized using one thread per  $x, y, (t)$  pixel (Supplementary Note 4). Our now entirely GPU-accelerated localization pipeline performs about tenfold faster than our CPU-only pipeline on a single computer (Supplementary Table 1). However, even parallelized over multiple workers, a single computer was too slow to keep up with our imaging.

Leveraging our distributed data storage, we developed a task-distribution architecture that enables a multiprocessing ‘cluster of one’ on a single computer as well as multi-computer clusters. Distributing tasks with a preference to assign jobs to computers where the data are saved allows us to minimize network overhead within the cluster. Critically, the performance of our architecture scales approximately linearly, allowing tuning of localization speed simply by adding more

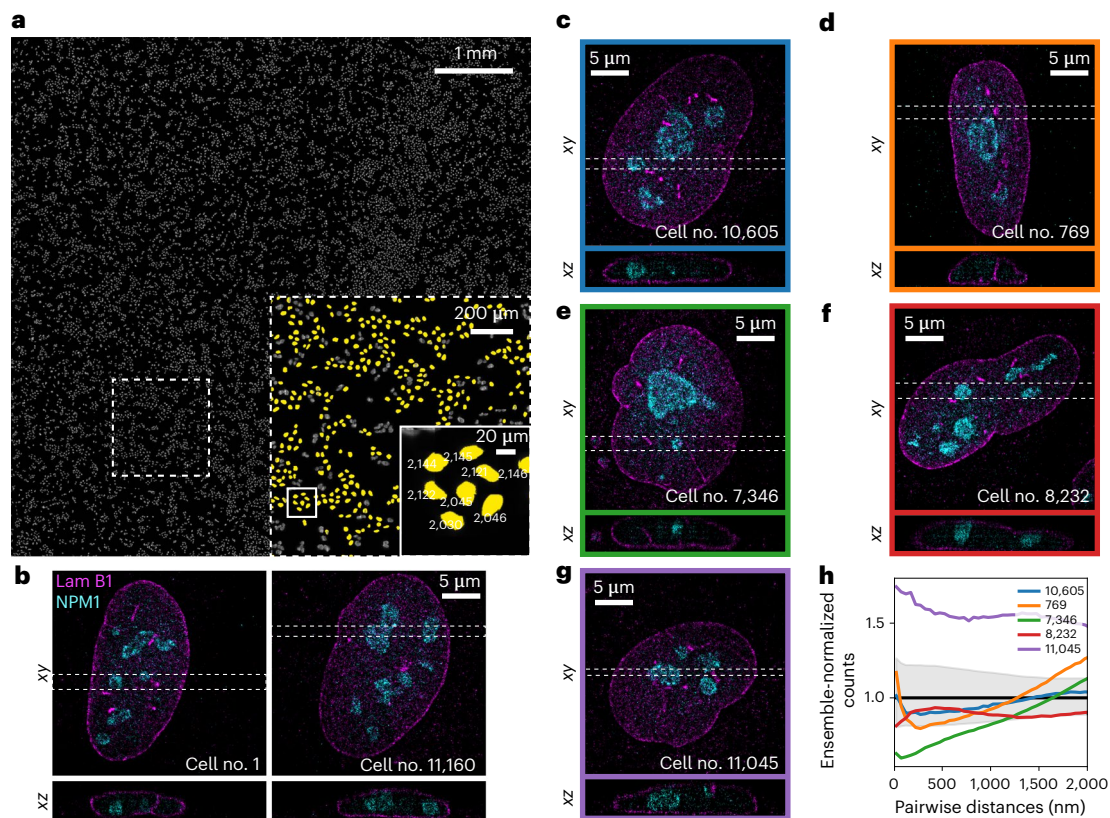
computers. Our production cluster consists of ten computers made from affordable consumer-grade components in 2016, each equipped with a GPU to run our accelerated algorithm (Supplementary Table 2). This additional factor of ~10 improvement in performance allows us to localize in real-time (Supplementary Table 1).

Localization tasks are automatically posted to the cluster on completion of recording or continually posted live during series acquisition for live visualization. A signal-to-noise-based candidate molecule detection threshold enables localization to be performed automatically across a wide range of conditions without user attention.

## 3D multicolor SMLM at 10,000 cells a day

The combination of our hardware and analysis advances enables us to image not only individual FOVs at high speed as shown in Fig. 2, but also 10,000 FOVs in a single day. To test this imaging mode, we plated U-2 OS cells on a coverslip and immunolabeled their nuclear lamini (anti-lamin b1, AF647) and nucleoli (anti-NPM1, CF568). Automated imaging begins by first scanning the coverslip in widefield mode and stitching together the images to create a large mosaic image. This image is then segmented and processed to generate a list of suitable FOVs for SMLM imaging. The overview mosaic shown in Fig. 4a was acquired in 52 min, after which automated super-resolution imaging of the 11,160 detected nuclei commenced. Each detected nucleus was imaged in 9.44 s, with the objective piezo actuator stepping over an axial range of 4.4  $\mu\text{m}$  for an axial localization range of approximately 5  $\mu\text{m}$ . The total imaging time for all 11,160 cells targeted was 1.2 d. Of the target FOVs, 99.8% were successfully acquired (Methods), resulting in a total of 3,589,123,170 fitted emitters from 75,431,069 raw frames (504,879,862 high-precision localizations after combining molecules present in





**Fig. 4 | 3D multicolor SMLM imaging of 10,000 cells a day.** **a**, Overview mosaic image of U-2 OS cells on a coverslip, from which 11,160 FOVs were automatically detected for imaging. A magnified view of the dashed box in the overview image is shown in the large inset, with a further magnification shown inside the solid box. Nuclei that were queued and imaged are highlighted in yellow and their queue number is displayed in the smaller inset. Each detected nucleus was automatically imaged, averaging 9.44 s per FOV, or 10,000 cells per 26.2 h. **b**, The first and last nuclei imaged. PCA on the SMLM datasets was used to select

representative cells, with **c**, blue, showing the cell whose principal components were closest to the mean (most typical). **d** and **e** show cells at  $+2$  and  $-2$  MADs, respectively from the mean along the first principal component. **f** and **g** show  $\pm 2$  MAD, respectively along the second principal component. **h**, Sum-normalized nucleophosmin–nucleophosmin pairwise distance histograms for these selected cells in **c–g** are shown after a secondary normalization to the ensemble median (black, interquartile range shown in gray).

more than one of the biplane views and filtering; Supplementary Fig. 8). The first and last nuclei imaged are shown in Fig. 4b. While they look relatively similar to each other, we also observed oddly shaped nuclei and nuclei that one could reasonably think were representative if only imaging a handful of cells (Supplementary Video 1). To demonstrate the latter, we performed a principal component analysis (PCA) on a collection of features extracted from the SMLM localization data (Supplementary Table 3 and Supplementary Fig. 9). We show the closest cell to the PCA-space mean position (Fig. 4c), and cells located 2 median absolute deviations (MADs) away along both of the principal axes (Fig. 4d–g). Figure 4h shows the ensemble median-normalized NPM1–NPM1 pairwise distances of these selected cells, which vary substantially outside the interquartile range depicted by the gray-shaded area.

### Flexible analysis and integrated workflows

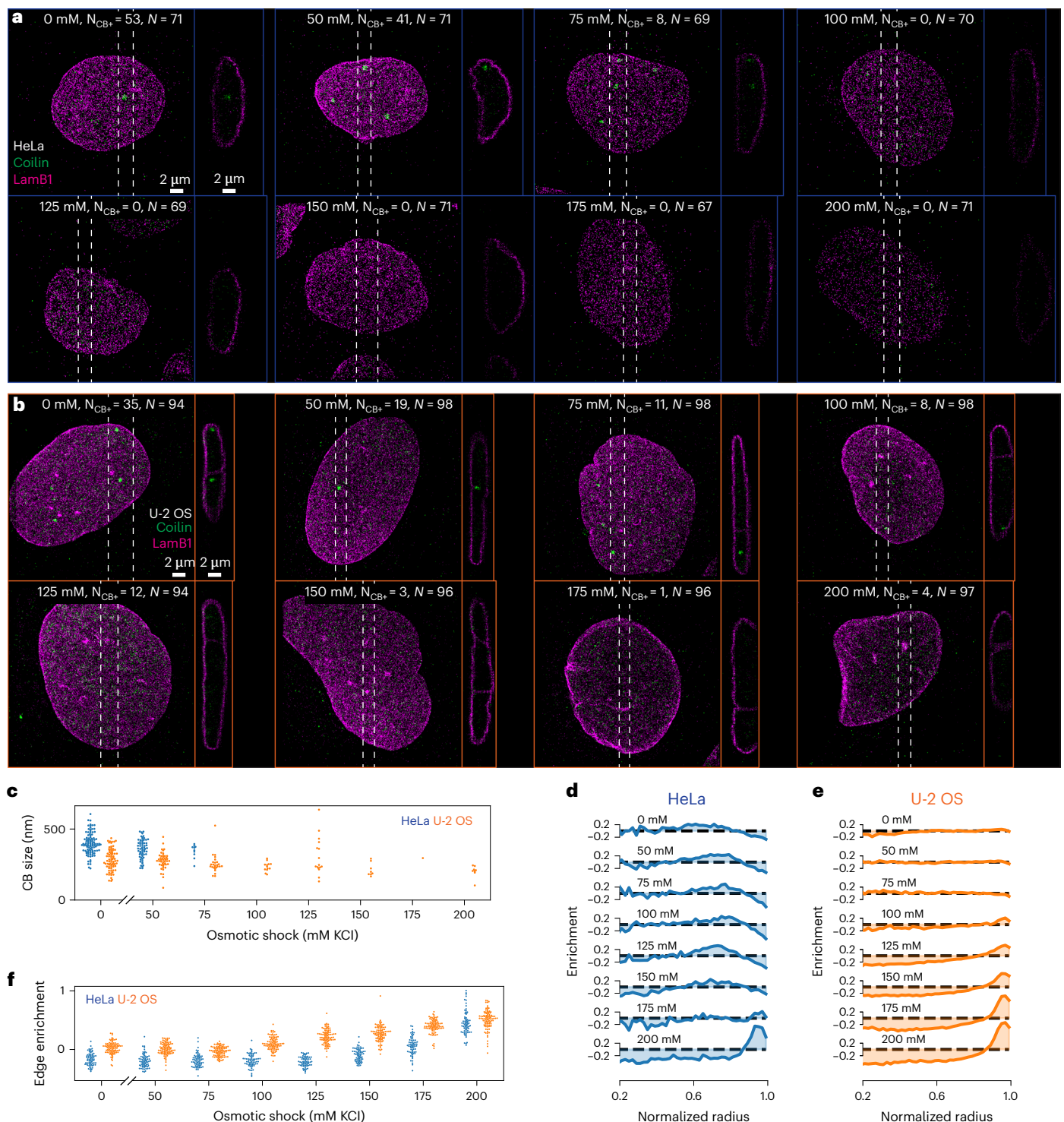
Analyzing thousands of SMLM localization datasets has to date been nontrivial, and often relied on patching together multiple existing packages. The analysis backbone of PYME, ‘recipes’, can be graphically reconfigured, can handle hybrid pointcloud/image-data workflows, and can be quickly built and tested in PYME’s interactive viewers<sup>14</sup>. Recipes can be efficiently batched on a single computer, or run on the cluster using the same analysis distribution we employ for localization tasks. They can additionally be chained together, allowing automatic localization and post-localization analyses workflows.

Microscope control solutions for custom systems are often bespoke, with their control flows hard-coded, making maintenance and

code sharing difficult. Many automated control flows consist of image analysis steps, particularly for more intelligent automation, which can quickly compound the amount of special-casing in instrument control code if these become at all sample-specific. PYMEAcquire, used here, allows intricate and easily reconfigurable workflows by using a priority queue for acquisition tasks. Acquisition protocols can be quickly written as lists of tasks to be executed on specified frames. Analysis chains can then be built graphically and linked to specific protocols, triggering (remote) execution on the PYME cluster at the beginning or end of a series acquisition as set by the user (Supplementary Video 2 and Supplementary Note 5). Finally, server endpoints in PYMEAcquire allow remote queuing of acquisition tasks, and can be leveraged in recipe modules to add acquisition tasks to the priority queue based on analyses of previously acquired series. This establishes a flexible architecture, fully integrating the instrument with distributed analysis.

### High-content screening with localization microscopy

Conducting entire multi-condition studies by SMLM has until now been impractical. Even in automated workflows, low frame rates coupled with acidifying glucose-oxidase/-catalase STORM buffers have limited the total number of ROIs imaged per plate to less than 100 (ref.<sup>8</sup>). We leveraged our platform to study the distribution and size of Cajal bodies (CBs), the site of small noncoding RNA transcription and processing<sup>21,22</sup>, in cells undergoing osmotic shock. Cellular stresses disrupt gene transcription, and as many nuclear bodies are hypothesized to be nucleated by transcription of specific genes, we expect the global



**Fig. 5 | Coilin redistributes in response to osmotic shock.** **a, b**, Example PCA-selected ROI from an osmotic shock experiment on HeLa cells (**a**) and U-2 OS cells (**b**) in automatically imaged eight-well slides. The concentration of KCl, number of ROIs containing a segmented CB ( $N_{CB+}$ ) and number of ROIs successfully imaged and analyzed ( $N$ ) are annotated on the example ROI for each condition.

**c**, Size of CBs at each concentration. **d, e**, Coilin enrichment relative to a uniform random distribution within a fitted nucleus model for HeLa cells (**d**) and U-2 OS cells (**e**). **f**, Coilin edge enrichment for each ROI, calculated as the average coilin enrichment relative to a uniform random distribution at normalized radii larger than 0.85.

disruption of transcription and RNA synthesis induced by osmotic shock to produce profound effects on CB integrity<sup>23</sup>. Because there is a large cell-to-cell variability in both the number and size of CBs, this work requires a substantial number of cellular images per condition to achieve a statistically robust comparison between osmotic shock conditions and greatly benefits from a high-throughput approach.

We imaged HeLa cells immunolabeled for coilin, a CB marker, with Cy3b, and lamin b1 with CF660C, in an eight-well slide format. These cells were shocked for 1 h before fixation using media supplemented with different concentrations of KCl in each well. A tile overview in the lamin channel was queued for each well, and we chained the overview protocol to a nucleus detection recipe to queue 45,000-frame 800-Hz



SMLM series at a maximum of 75 ROIs per well in a path-optimized manner and at a higher priority than the tile overview tasks such that large movements across the slide were minimized. We observed a relatively consistent presence of CBs per cell in the control condition (median, mean, s.d. = 1, 1.4, 1.1,  $N = 71$ ; see Supplementary Fig. 10 for a histogram of number of CBs per ROI). An ROI automatically selected as representative of each well by PCA is shown in Fig. 5a (Methods). The CBs were indeed profoundly affected, and were depleted in wells that had been subjected to substantial osmotic shock, with no CBs detectable at concentrations higher than 75 mM KCl. Additionally, the CBs decreased in size with increasing osmolarity (Fig. 5c).

To see if this was an idiosyncrasy of the cell type or a more universal phenomenon, we repeated the experiment in U-2 OS cells, imaging a maximum of 100 ROIs in each well for 33,000 frames each (Fig. 5b). Notably, the CBs in the U-2 OS control cells are smaller, and their number more varied (median, mean, s.d. = 0, 0.8, 1.4,  $N = 94$ ; see also histograms in Supplementary Fig. 10), making them generally more difficult to study at low throughputs. Further, while the osmotic shock still affects both CB presence and size, these effects are more subtle than in the HeLa cells (Fig. 5c).

While the dissipation of CBs under osmotic shock could be rigorously monitored with our system, the function of most proteins in membraneless organelles is not limited to these structures, and we noticed an opportunity to leverage our statistical power to measure the redistribution of coilin. For each ROI, we fit the lamin localizations to a spherical harmonic shell to define a nuclear coordinate system for the coilin localizations. We see coilin enrichment (relative to uniform distribution simulated in the shell) at the nuclear periphery in stressed HeLa and U-2 OS cells (Fig. 5d–f).

## Discussion

We have shown that it is possible to carry out large SMLM studies without sacrificing sample population or experimental conditions. This is a substantial departure from the current state-of-the-art. The software advances this required are not only enabling of higher-throughput studies, but should additionally reduce frustrations common with smaller experiments or even manual exploration. The modular and plugin-friendly nature of PYME and its interactive data viewers make it feasible for users to easily explore and extract user-defined features from these large datasets. Similarly, PYME is well suited for instrument developers to easily extend their acquisition or analysis capabilities in a complete environment.

Automated SMLM at high bandwidths enables users to acquire vast datasets at raw frame rates faster than they can visually assess, which warrants consideration of tools to assist them. For example, localization quality control such as Wasserstein-induced Flux<sup>24</sup> and ROI error-mapping methods<sup>25,26</sup> could be helpful in determining if imaging and localization analysis were performed adequately. PYME recipes can produce HTML report outputs. Plugins to calculate such measures of quality control could facilitate automatic checks for each series or acquisition run, which can be formatted as HTML reports within recipes.

In addition to monitoring image quality, our approach could be improved by further optimizing automated acquisition, for example, regulating emitter density by servo-controlled laser intensities, varying the number of frames acquired at a given ROI depending on its actual sampling requirements or correcting sample-induced aberrations at each ROI. While impressive progress has been made in the area of intelligent SMLM automation<sup>27</sup>, there are many unmet challenges to apply these advances to 3D and/or multicolor imaging, and especially for high-speed imaging at camera frame rates of 800 Hz.

We expect hardware advances such as higher-bandwidth cameras with reduced amplification noise and larger FOVs to continue to improve SMLM image quality and capabilities. We additionally note that per-pixel quantization using noise envelope scaling as demonstrated

in this work could be performed on-board cameras, and could aid in further bandwidth optimizations. Integrating dense-emitter localization algorithms such as 3D-DAOSTORM<sup>28</sup> or DECODE<sup>29</sup> could reduce the number of raw frames required for a reconstruction. For structures that are highly stereotypical and easily inferred from sparse or low-resolution images, structure-prediction methods such as ANNA-PALM (artificial neural network accelerated PALM)<sup>30</sup> could also be incorporated within our framework to further improve throughput.

With the advances contained in this work, we expect SMLM to become a much more routinely and broadly applicable tool. The example applications in this work are only a small sampling of the possibilities that automated high-throughput SMLM offers, but demonstrate the technical feasibility both of obtaining nano-scale resolution in every cell of an extended specimen and of combining this with the systematic investigation of multiple conditions. Exciting potential applications of our work include looking at the cell-type-dependent distribution of protein clustering within tissue sections, the detection of statistically rare but functionally important configurations (for example, transient contacts between gene regions) and the use of SMLM as a screening technique to assess structural responses to a wide range of interventions. These types of experiments will improve understanding of disease in ways that were previously not possible.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-023-01702-1>.

## References

1. Baddeley, D. & Bewersdorf, J. Biological insight from super-resolution microscopy: what we can learn from localization-based images. *Annu. Rev. Biochem.* **87**, 965–989 (2018).
2. Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456 (2013).
3. Szymborska, A. et al. Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341**, 655–658 (2013).
4. Zhang, Y., Lara-Tejero, M., Bewersdorf, J. & Galán, J. E. Visualization and characterization of individual type III protein secretion machines in live bacteria. *Proc. Natl Acad. Sci. USA* **114**, 6098–6103 (2017).
5. Yuan, P. et al. Trem2 haploinsufficiency in mice and humans impairs the microglia barrier function leading to decreased amyloid compaction and severe axonal dystrophy. *Neuron* **90**, 724–739 (2016).
6. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
7. Mund, M. et al. Systematic nanoscale analysis of endocytosis links efficient vesicle formation to patterned actin nucleation. *Cell* **174**, 884–896.e17 (2018).
8. Beghin, A. et al. Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* **14**, 1184–1190 (2017).
9. Holden, S. J. et al. High throughput 3D super-resolution microscopy reveals *Caulobacter crescentus* in vivo Z-ring organization. *Proc. Natl Acad. Sci. USA* **111**, 4566–4571 (2014).
10. Huang, F. et al. Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10**, 653–658 (2013).
11. Lin, Y. et al. Quantifying and optimizing single-molecule switching nanoscopy at high speeds. *PLoS ONE* **10**, e0128135 (2015).

12. Diekmann, R. et al. Optimizing imaging speed and excitation intensity for single-molecule localization microscopy. *Nat. Methods* **17**, 909–912 (2020).
13. Shim, S.-H. et al. Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes. *Proc. Natl Acad. Sci. USA* **109**, 13978–13983 (2012).
14. Marin, Z. et al. PYMEVisualize: an open-source tool for exploring 3D super-resolution data. *Nat. Methods* **18**, 582–584 (2021).
15. Juette, M. F. et al. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nat. Methods* **5**, 527–529 (2008).
16. Olivier, N., Keller, D., Gönczy, P. & Manley, S. Resolution doubling in 3D-STORM imaging through improved buffers. *PLoS ONE* **8**, 1–9 (2013).
17. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
18. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
19. Lin, R., Clowsley, A. H., Jayasinghe, I. D., Baddeley, D. & Soeller, C. Algorithmic corrections for localization microscopy with sCMOS cameras—characterisation of a computationally efficient localization approach. *Opt. Express* **25**, 11701 (2017).
20. Smith, C. S., Joseph, N., Rieger, B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods* **7**, 373–375 (2010).
21. Staněk, D. & Neugebauer, K. M. Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. *J. Cell Biol.* **166**, 1015–1025 (2004).
22. Machyna, M. et al. The coilin interactome identifies hundreds of small noncoding RNAs that traffic through Cajal bodies. *Mol. Cell* **56**, 389–399 (2014).
23. Boulon, S., Westman, B. J., Hutten, S., Boisvert, F. M. & Lamond, A. I. The nucleolus under stress. *Mol. Cell* **40**, 216–227 (2010).
24. Mazidi, H., Ding, T., Nehorai, A. & Lew, M. D. Measuring localization confidence for quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy. In *Proc. SPIE 11246, Single Molecule Spectroscopy and Superresolution Imaging XIII* (eds Gregor, I. et al.) 1124611 (SPIE, 2020).
25. Culley, S. et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* **15**, 263–266 (2018).
26. Marsh, R. J. et al. Sub-diffraction error mapping for localisation microscopy images. *Nat. Commun.* **12**, 1–13 (2021).
27. Štefko, M., Ottino, B., Douglass, K. M. & Manley, S. Autonomous illumination control for localization microscopy. *Opt. Express* **26**, 30882 (2018).
28. Babcock, H., Sigal, Y. M. & Zhuang, X. A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. *Opt. Nanoscopy* **1**, 6 (2012).
29. Speiser, A. et al. Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat. Methods* **18**, 1082–1090 (2021).
30. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023



## Methods

### High-throughput SMLM microscope

The optical setup, shown in Fig. 1b, is detailed in Supplementary Note 1.1 and Supplementary Fig. 1, and discussed here only briefly. The microscope hardware differs from conventional SMLM microscopes in that an sCMOS camera and two high-power (2 W) excitation lasers (560 nm and 642 nm) are used. A cylindrical lens to provide astigmatism is combined with a biplane module to extend the axial localization range. The laser lines are coupled into a multimode fiber, which is vibrated to average out speckle and achieve a uniform intensity profile at the fiber exit which is imaged into the sample. A fast objective piezo actuator enables axial scanning, and a motorized sample stage accommodates lateral sample movement on the order of 100 mm. The microscope additionally has a custom-built focus lock, where a near-infrared laser is reflected off the coverslip at an angle such that the reflected beam position can be monitored by an additional camera. This position indicates the distance between the objective and the coverslip, and can be servo-controlled.

The instrument was originally controlled with a custom LabView program (Phase 1), which was used for imaging in Figs. 2 and 4, and would save DCIMG-formatted image files to a RAM disk where a PYME script would open them and spool them to the computer cluster, additionally launching localization analysis. We then converted instrument control to PYMEAcquire for further integration of the system (Phase 2), after which we no longer saved frames to disk on the instrument computer as we could compress and spool them directly.

### SMLM imaging

Alexa Fluor 647 (AF647) and CF660C are suitable for high-speed SMLM<sup>11,12</sup>, CF660C being particularly robust. We additionally found CF568 to be capable of fast switching, enabling high-speed multicolor SMLM in standard STORM buffers using the glucose-oxidase/-catalase oxygen scavenging system ('glox' buffer). Not all imaging formats are easily sealed, and we find sulfite-based oxygen scavenging in aqueous buffer<sup>31</sup> performs well at high speeds, and is compatible with Cy3B and CF660C, again enabling sustained high-speed multicolor imaging.

The glox buffer used in this work consisted of 0.49 kU ml<sup>-1</sup> glucose oxidase, 0.98 kU ml<sup>-1</sup> glucose catalase, 0.14 M 2-mercaptoethanol, 2.5 mM Tris at pH 8, 2.5 mM NaCl and 0.28 M glucose in milli-q water. The aqueous sulfite buffer used in this work consisted of 20 mM Na<sub>2</sub>SO<sub>3</sub> and 0.14 M 2-mercaptoethanol in 1 × PBS.

**Sealed glox buffer imaging.** All samples in Figs. 2 and 4 were imaged in the glox STORM buffer. These samples were mounted and imaged in Biopetech FCS2 flow chambers with the microaqueduct slide flipped upside down such that the flow input/output tubes were sealed. Before mounting the sample, the edges were scraped with a scalpel to remove cells that would interfere with ideal sealing. We used 0.5-mm-thick circular gaskets, overfilled the chamber with 500 µl of STORM buffer and rinsed the coverslip in 200 µl of STORM buffer before mounting. Once the coverslip was in place, it was flattened (and excess buffer was removed) by pressing a wipe against it using a flat plastic dish.

All SMLM imaging shown in Figs. 2 to 4 was performed at a camera frame rate of 800 Hz. The 560-nm and 642-nm laser intensities delivered to the samples were approximately 71 kW cm<sup>-2</sup> and 51 kW cm<sup>-2</sup> for the lamin B1 and NPM1 samples, 41 kW cm<sup>-2</sup> and 51 kW cm<sup>-2</sup> for the lamin A/C and lamin-associated chromatin domain (LAD) pool 2 samples, 61 kW cm<sup>-2</sup> and 51 kW cm<sup>-2</sup> for the α-tubulin and Sec61β samples, 38 kW cm<sup>-2</sup> and 51 kW cm<sup>-2</sup> for the double-stranded DNA (dsDNA) and TOM20 samples, and 41 kW cm<sup>-2</sup> and 61 kW cm<sup>-2</sup> for the chromosome 22 topologically associating chromatin domains (TADs) and LAD pool 1, respectively. The 405-nm activation light was not used during SMLM data acquisition of these samples.

The 3D images of the α-tubulin and Sec61β samples, and the TOM20 and dsDNA samples, were acquired without 'z-stepping' during

recording. The remaining samples in Figs. 2 and 4 were imaged with two or six cycles of seven interwoven 'up then down' z-steps to obtain even axial localization distribution over the thicker volumes.

**Sulfite buffer imaging.** The HeLa and U-2 OS samples in Fig. 5 were imaged in an aqueous sulfite STORM buffer, which enables flexible sample formats such as small petri dishes, eight-well slides and multi-well plates. For these samples, N<sub>2</sub> was additionally perfused into a stagetop incubator chamber (OKO K-Frame) at a rate of 0.2 l min<sup>-1</sup>. Stage leveling was checked before imaging, and the offsets measured by the leveling routine were stored to bootstrap automatically finding the coverslip should the focus lock ever be lost.

All SMLM imaging with the sulfite buffer (the coilin and lamin samples in Fig. 5) was performed at 783 Hz. The 560-nm and 642-nm intensities delivered to the samples were approximately 49 kW cm<sup>-2</sup> and 53 kW cm<sup>-2</sup>. Imaging began on each ROI only with 642-nm illumination, turning on the 560-nm laser after a full z-stack cycle (18,000 frames for the 45,000-frame HeLa series, and 11,000 frames for the 33,000 U-2 OS series). Additionally, <0.08 kW cm<sup>-2</sup> of 405-nm illumination was applied nearer the end of each series. The U-2 OS series of 33,000 frames were acquired in about 44 s each.

### Compression

Before spooling the data from the instrument computer RAM to the computer cluster, we minimize the data volume via compression. This compression includes three steps: offset subtraction, re-quantization such that the quantization intervals scale with the square root of the number of photons and then Huffman coding<sup>32</sup>. The re-quantization step serves to decrease the number of discrete levels within the data, dramatically increasing the efficiency of Huffman coding. This step at each/pixel can be expressed as

$$d_j = \lfloor \frac{\eta}{q} \sqrt{D_j - o_j} \rfloor, \quad (1)$$

where  $D$  is the intensity counts in a single frame,  $o$  is the camera offset map,  $d$  is the resulting quantized intensity bin,  $\eta$  is the number of photoelectrons per ADU and  $q$  is the desired scale factor in Poisson noise units (that is, 0.5 for  $\sigma/2$  quantization).  $\lfloor x \rfloor$  denotes a rounding operation on  $x$  to the nearest integer.

Our approach achieves a 5–8-fold reduction in data volume. While the quantization step is lossy, we choose the quantization intervals such that the additional quantization noise introduced is less than the Poisson noise we would expect from photon detection and therefore has a negligible influence on subsequent processing and localization. Our approach is similar to one developed separately and reported in ref.<sup>33</sup>. Our optimized C code leverages the AVX SIMD (single instruction multiple data) instruction set when available and allows the compression to run at ~1 GB s<sup>-1</sup> without the need for specialized hardware (for example, GPUs) on the acquisition machine (Supplementary Fig. 4).

### Lamin and nucleophosmin PCA

Of the 11,160 series imaged, 11,136 were spooled and localized during the same time, with analysis being run post-facto on an additional 9 series for a total of 11,145 SMLM images. Spooling failed for 15 images. Additionally, pointclouds with fewer than 5,000 localizations were ignored. For each of the remaining 11,117 lamin and nucleophosmin SMLM pointclouds, 13 metrics were calculated and used to create a feature vector. These metrics are calculated directly from the localizations and are described in Supplementary Table 3. The feature vector was normalized by the interquartile range along each metric, and this 11,117 × 13 array was then reduced to size 11,117 × 2 using PCA limited to two components. The features for each of these 11,117 pointclouds were projected onto the resulting principal components 0 and 1 (PC0, PC1) two-dimensional (2D) basis and are plotted in Supplementary

Fig. 9. The MAD was calculated along PC0 and PC1, as was the mean of the ensemble. A KDTree of PC0/PC1 Euclidean distance was generated and the cells nearest to the mean and the mean  $\pm 2$  MAD along each axis were selected as shown in Supplementary Fig. 9, and as displayed in Fig. 4c.

### CB measurements and coilin distribution analysis

CBs were segmented from coilin localizations using a Bayesian information criterion (BIC)-optimized Gaussian mixture, and lamin localizations were fitted to an expansion of spherical harmonics to create a radial reference frame for each coilin localization.

First, a generous 2D binary mask of the nucleus was used to isolate signal localization for all further analyses. This mask was generated from the lamin localizations by rendering a 2D  $\sigma = 30$ -nm Gaussian image with 40-nm pixels, and applying a  $15 \times 15$ -pixel maximum filter, followed by an Otsu threshold. Regions other than the largest were discarded, and the nucleus mask was dilated by 1 pixel 10 times.

**Radial coilin distribution analysis.** The lamin localizations were least squares fitted to an expansion of 16 spherical harmonic functions (the  $l \in \{0, 1, 2, 3\}$  modes) centered at the lamin localization center of mass, producing an analytic shell representation of the nuclear envelope. Coilin localizations were converted from Cartesian to spherical coordinates, again relative to the lamin center of mass, and the shell radius at their zenith and azimuth coordinate was used to normalize their radial coordinate. This shell-normalized coilin distribution was histogrammed into 50 bins and sum-normalized. To compare the coilin distribution with a uniform distribution, this process was repeated for a uniform-density pointcloud simulation bounded by the same shell (5 iterations targeting a density of 1 point per isotropic 200-nm voxel). Finally, a per-nucleus coilin-enrichment histogram was created by dividing the shell-normalized coilin radius histogram by that of the uniform-density simulation and subtracting 1. The per-well average of this coilin-enrichment histogram is shown in Fig. 5d,e. The per-nucleus edge enrichment (Fig. 5f) was calculated by taking the average value of the coilin enrichment in radial bins above 0.85.

**CB measurements.** A Gaussian mixture model (GMM) was fitted to the coilin localizations. For each ROI, the number of components was stepped by 1 from 1 to 50, and the BIC was calculated for each. Metrics were calculated for each candidate CB as shown in Supplementary Table 4. These candidate CBs were then filtered, rejecting candidates with fewer than 20 localizations, greater than 711-nm MAD (corresponding to a Gaussian FWHM of  $\sim 2.5 \mu\text{m}$ ) and an approximate density of at least  $5 \times 10^{-6} \text{ nm}^{-3}$  (with volume calculated as  $v = \frac{4}{3}\pi r_g^3$ , where  $r_g$  is the gyration radius).

The ROIs shown in Fig. 5a,b were selected using PCAs of these per-CB features by well. For each well, if there were CBs in less than 5% of the ROIs, the feature vector (Supplementary Table 4) was constructed from the raw BIC-optimized GMM results (candidate CBs) rather than the filtered (CBs), as was otherwise used. After that decision, the feature vector was scaled by centering on the feature medians and normalizing by their interquartile ranges before performing a three-component PCA. The feature vector was then projected onto these principal components, and the median value along each component was used to query a KDTree of PC0/PC1/PC2 Euclidean distances. Each ROI shown in Fig. 5 contains either the closest or second closest CB candidate to that point.

The GMM implementation used is in Scikit-Learn<sup>34</sup>, and each component was fitted with its own general covariance matrix.

### Reporting summary

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

### Data availability

The 3D localizations, calibration files and raw blinking videos for all series in Fig. 2, and Cell no. 1, no. 2,504, no. 5,735, no. 8,041, no. 9,577 and no. 11,160 from the lamin-NPM1 dataset in Fig. 4 (and 3D localizations for the remaining cells), are publicly available through the 4D Nucleome data portal at <https://data.4dnucleome.org/publications/7d9fad19-54c4-419e-8d99-8157f5c1904b/>. Any additional data from this work can be obtained through the authors upon request.

### Code availability

The code for automated acquisition, distributed data storage and analysis is released under the GNU General Public License v.3 as part of the python-microscopy project and is available at [github.com/python-microscopy/python-microscopy](https://github.com/python-microscopy/python-microscopy). The quantized compression software can be installed independently, with instructions for use with third-party software additionally available at [github.com/python-microscopy/pymecompress](https://github.com/python-microscopy/pymecompress). Code for GPU acceleration of single-molecule fitting is available under an academic use license from [github.com/barentine/pyme-warp-drive](https://github.com/barentine/pyme-warp-drive). The LabVIEW acquisition software used in phase 1 can be obtained from the authors; however, it is not actively maintained. Please contact the authors for alternative licensing arrangements.

### References

- Hartwich, T. M. et al. A stable, high refractive index, switching buffer for super-resolution imaging. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/465492v1> (2018).
- Huffman, D. A. A method for the construction of minimum-redundancy codes. *Proc. IRE* **40**, 1098–1101 (1952).
- Balazs, B., Deschamps, J., Albert, M., Ries, J. & Hufnagel, L. A real-time compression library for microscopy images. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/164624v1> (2017).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

### Acknowledgements

We thank A. Wrogg for contributing the timeline in Fig. 1. We thank L. Schroeder and Y. Zhang for helpful discussions and technical assistance. This work was primarily supported by a 4D Nucleome grant from the National Institutes of Health (NIH) (grant no. U01 DA047734 to J.B. and D.B.). J.B. acknowledges support from NIH grant no. P30 DK045735 (to R. Sherwin). A.E.S.B. acknowledges support by an NIH training grant (no. T32 GM008283) and training on the Computational Image Analysis in Cellular and Developmental Biology Course of the Marine Biology Laboratory (which was supported by NIH grant no. R25 GM103792). We are also grateful for funding from NIH awards no. U01CA200147 TPCA-2017-Neugebauer and OPAS (to K.M.N.) and no. 1R01NS128358-1 (to K.M.N. and J.B.). This work is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Author contributions

Y.L. and J.B. designed the optical hardware of the microscope which Y.L. built. A.E.S.B., Y.L., D.B., Z.M., T.P. and J.R.C. developed acquisition control software. M.R.G., A.E.S.B. and D.B. designed and implemented the computer cluster. D.B. designed the distributed storage architecture and compression algorithm. D.B. and L.B. designed and implemented the cluster task distribution. A.E.S.B. and D.B. developed the GPU acceleration code. S.W. and M. Liu designed the FISH probes. E.M.C., P.K., M. Lessard, F.R.-M., M. Liu, S.W. and K.M.N. optimized sample preparation protocols and prepared samples. A.E.S.B., Y.L. and E.M.C. performed imaging experiments. A.E.S.B., Y.L. and D.B. performed post-localization analysis. All authors contributed to writing the manuscript.

## Competing interests

J.B. discloses a financial interest in Bruker Corp. and Hamamatsu Photonics. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-023-01702-1>.

**Correspondence and requests for materials** should be addressed to Joerg Bewersdorf or David Baddeley.

**Peer review information** *Nature Biotechnology* thanks Bogdan Bintu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), 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 | <p>Data collection was performed using custom LabView(TM) and Python code.</p> <p>The code for automated acquisition, distributed data storage and analysis is released under the GNU General Public License v3 as part of the python-microscopy project and is available at <a href="#">github.com/python-microscopy/python-microscopy</a>. The quantized compression software can be installed independently, with instructions for use with third-party software available at <a href="#">github.com/python-microscopy/pymecompress</a>.</p> <p>Code for GPU acceleration of single-molecule fitting is available under an academic use license from <a href="#">github.com/barentine/pyme-warp-drive</a>.</p> <p>Though the course of the project we used multiple versions of python (2.7, 3.6, 3.7, 3.8). The code in it's current state is well tested on Python 3.7 and 3.8.</p> <p>The LabVIEW acquisition software used in phase 1 can be obtained on request from the authors, but is no longer actively maintained. Please contact the authors for alternative licensing arrangements.</p> <p>OligoArray 2.1 and BLASTN 2.2.30 were used for designing LAD fish sequences.</p> |
| Data analysis   | <p>Data analysis was performed using the same custom python packages as above (python, python-microscopy, pymecompress, pyme-warp-drive).</p>  |

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Due to data volume it was not possible to store all out experimental data on public repositories. A representative subset, namely the 3D localizations, calibration files, and raw blinking movies for all series in Figure 2, and Cell #1, #2504, #5735, #8041, #9577, and #11,160 from the lamin-NPM1 dataset in Figure 4 (and 3D localizations for the remaining cells) are publicly available through the 4D Nucleome data portal at <https://data.4dnucleome.org/publications/7d9fad19-54c4-419e-8d99-8157f5c1904b/>. Any additional data from this work can be obtained through the authors upon reasonable request.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

|                             |     |
|-----------------------------|-----|
| Reporting on sex and gender | N/A |
| Population characteristics  | N/A |
| Recruitment                 | N/A |
| Ethics oversight            | N/A |

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

## 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](https://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     | For the Lamin-Nucleophosmin dataset, sample size (11,160 cells) was chosen to demonstrate that we could image a large population of cells within a single experiment - sufficient to explore population effects and to detect rare configurations.<br><br>For the Coilin-Lamin data sets, sample size (100 cells /condition)   |
| Data exclusions | Of the 11,160 fields of view comprising the Lamin B1 and NPM1 dataset, 23 fields of view were excluded due to failed spooling or automatic quality control, as described in the Methods.<br><br>For the Coilin-Lamin data set in U2OS, we excluded 24 of 795 fields of view on the basis of automatic quality control, as described in Methods.<br><br>For the Coilin-Lamin data set in HeLa, we manually excluded 70 fields of view which were inadvertently imaged twice, and an additional 44 fields of view on the basis of automatic quality control, from a total of 673 acquisitions. |
| Replication     | As the focus of this paper is on a technical demonstration, no replicates were performed for the Lamin and Nucleophosmin dataset. No technical replicates were performed for the Coilin-lamin data set, although we did repeat the experiment in HeLa cells to show that the effect is not limited to U2OS.  |
| Randomization   | The Lamin-nucleophosmin data consisted of only one group, so randomization was not applicable.<br><br>Within the Coilin-Lamin sets, all osmotic shock concentrations (groups) were imaged in one experimental run and were interleaved in the imaging order to ensure that imaging time did not correlate with osmotic shock. This mitigated the potential for any buffer-rundown effects on image quality to effect results.  |
| Blinding        | The Lamin-nucleophosmin dataset did not have allocated groups, making blinding not applicable.<br>For Coilin-Lamin data sets, analysis parameters (clustering thresholds) were determined using a couple of randomly selected cells in the   |

control (no osmotic shock) group, and the same parameters applied automatically to the analysis of all other cells, removing the opportunity for bias in the analysis.

## 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    |   | Methods                             |   |
|-------------------------------------|---|-------------------------------------|---|
| n/a                                 | Involved in the study                                     | n/a                                 | Involved in the study                           |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies            | <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Eukaryotic cell lines | <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology    | <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms      |                                     |   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                    |                                     |   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern     |                                     |   |

## Antibodies

### Antibodies used

mouse anti-tubulin (T5168, Sigma-Aldrich)  
 rabbit anti-GFP (A-11122, Invitrogen)  
 mouse anti-nucleophosmin (NB600-1030, Novus Biologicals)  
 rabbit anti-lamin b1 (ab16048, Abcam)  
 mouse anti-dsDNA (ab27156, Abcam)  
 rabbit anti-TOM20 (ab78547, Abcam)  
 mouse anti-lamin A/C (4C11, Cell Signaling Technology)

goat anti-rabbit CF568 (20099, Biotium)  
 goat anti-mouse CF568 (20109, Biotium)  
 goat anti-mouse CF568ST (20800, Biotium)  
 goat anti-rabbit AF647 (A21245, Invitrogen)  
 goat anti-mouse AF647 (A21237, Invitrogen)  
 goat anti-mouse Biotin (115-065-003, Jackson ImmunoResearch Laboratories)

### Validation

The results of all primary antibodies are consistent with information provided by the manufacturer. Beyond their manufacturer validation measures, we verified sub-cellular localisation was as expected in immunofluorescence:

Mouse anti-tubulin (T5168, Sigma-Aldrich), validated by manufacturer with immunofluorescence and immunoblot  
 Rabbit anti-GFP (A-11122, Invitrogen), validated by manufacturer using immunofluorescence and western blotting of multiple GFP-conjugated expressed proteins.  
 Mouse anti-nucleophosmin (NB600-1030, Novus Biologicals), validated by manufacturer using western blot, immunofluorescence, and immunoblot  
 Rabbit anti-lamin b1 (ab16048, Abcam), validated by manufacturer using western blot, immunofluorescence, and immunohistochemistry  
 Mouse anti-dsDNA (ab27156, Abcam), validated by manufacturer using immunohistochemistry and dot blots  
 Rabbit anti-TOM20 (ab78547, Abcam), validated by manufacturer using immunohistochemistry, immunofluorescence, western blot and immuno-precipitation.  
 Mouse anti-lamin A/C (4C11, Cell Signaling Technology), validated by manufacturer using western blotting, immunohistochemistry, and immunofluorescence  
 Mouse anti-coilin (ab87913, Abcam), validated by manufacturer using western blotting, immunohistochemistry, immunofluorescence, immunoprecipitation, and flow cytometry.

All secondary antibodies are validated by immunofluorescence in this work.

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

### Cell line source(s)

IMR-90 ATCC CCL-186  
 U-2 OS ATCC HTB-96  
 COS-7 ATCC batch #63624240  
 HeLa-Kyoto, Neugebauer Lab RRID:CVCL\_1922

### Authentication

IMR-90: purchased directly from ATCC, no further authentication was performed  
 U-2 OS: purchased directly from ATCC, no further authentication was performed  
 COS-7: purchased directly from ATCC, no further authentication was performed  
 HeLa-Kyoto: no specific authentication was performed



Mycoplasma contamination

Cell lines were not tested for Mycoplasma contamination

Commonly misidentified lines  
(See [ICLAC](#) register)

N/A