

# Performance of deep learning restoration methods for the extraction of particle dynamics in noisy microscopy image sequences

Paul Kefer<sup>a,†</sup>, Fadil Iqbal<sup>b,†</sup>, Maelle Locatelli<sup>c</sup>, Josh Lawrimore<sup>d</sup>, Mengdi Zhang<sup>b,e</sup>, Kerry Bloom<sup>d</sup>, Keith Bonin<sup>a,f</sup>, Pierre-Alexandre Vidi<sup>c,f,\*</sup>, and Jing Liu<sup>b,\*</sup>

<sup>a</sup>Department of Physics, Wake Forest University, Winston-Salem, NC 27109; <sup>b</sup>Department of Physics, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202; <sup>c</sup>Department of Cancer Biology, Wake Forest School of Medicine, and; <sup>d</sup>Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; <sup>e</sup>Department of Pharmacology, School of Pharmacy, Harbin Medical University, Harbin 150081, China; <sup>f</sup>Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC 27157

**ABSTRACT** Particle tracking in living systems requires low light exposure and short exposure times to avoid phototoxicity and photobleaching and to fully capture particle motion with high-speed imaging. Low-excitation light comes at the expense of tracking accuracy. Image restoration methods based on deep learning dramatically improve the signal-to-noise ratio in low-exposure data sets, qualitatively improving the images. However, it is not clear whether images generated by these methods yield accurate quantitative measurements such as diffusion parameters in (single) particle tracking experiments. Here, we evaluate the performance of two popular deep learning denoising software packages for particle tracking, using synthetic data sets and movies of diffusing chromatin as biological examples. With synthetic data, both supervised and unsupervised deep learning restored particle motions with high accuracy in two-dimensional data sets, whereas artifacts were introduced by the denoisers in three-dimensional data sets. Experimentally, we found that, while both supervised and unsupervised approaches improved tracking results compared with the original noisy images, supervised learning generally outperformed the unsupervised approach. We find that nicer-looking image sequences are not synonymous with more precise tracking results and highlight that deep learning algorithms can produce deceiving artifacts with extremely noisy images. Finally, we address the challenge of selecting parameters to train convolutional neural networks by implementing a frugal Bayesian optimizer that rapidly explores multidimensional parameter spaces, identifying networks yielding optimal particle tracking accuracy. Our study provides quantitative outcome measures of image restoration using deep learning. We anticipate broad application of this approach to critically evaluate artificial intelligence solutions for quantitative microscopy.

## Monitoring Editor

Alexander Mogilner  
New York University

Received: Nov 3, 2020

Revised: Jan 6, 2021

Accepted: Jan 13, 2021

This article was published online ahead of print in MBoc in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E20-11-0689>) on January 27, 2021.

<sup>†</sup>Equal contributions.

Data availability: Code for generation of synthetic data, localization analysis, and CNN parameter optimization can be accessed on GitHub (<https://github.com/LiuLabiUPUI/CARE-SMT>). The experimental data and simulations used in this article are available on the Open Science Framework (OSF; <https://osf.io/sehwb/osf.io>).

\*Address correspondence to: Pierre-Alexandre Vidi ([pvidi@wakehealth.edu](mailto:pvidi@wakehealth.edu)); or Jing Liu ([urralliu@iupui.edu](mailto:urralliu@iupui.edu)).

Abbreviations used: BM3D, block matching and 3-dimensional filtering; CARE, content aware image restoration; CCD, charged coupled device; CNN, convolu-

tional neural network; DN, denoised; GPU, graphical processing unit; GT, ground truth; MSD, mean squared displacement; N2V, noise to void; a deep learning-based denoising software; PAGFP, photoactivatable green fluorescent protein; sCMOS, scientific complementary metal–oxide–semiconductor; SMT, Single-Molecule Tracking; SNR, Signal to Noise Ratio; structN2V, structured Noise to Void.

© 2021 Kefer, Iqbal, et al. This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®,” “The American Society for Cell Biology®,” and “Molecular Biology of the Cell®” are registered trademarks of The American Society for Cell Biology.

## INTRODUCTION

Sample illumination is a conundrum in live-cell imaging: too much light leads to fluorophore photobleaching and phototoxicity (Carlton *et al.*, 2010; Magidson and Khodjakov, 2013), whereas too little light results in poor images with low signal-to-noise ratio (SNR). Low SNR is detrimental to quantitative microscopy applications requiring precise localization and tracking of particles (Chenouard *et al.*, 2014). The issue of illumination versus SNR is compounded in single-molecule tracking (SMT) experiments because each fluorescent molecule emits a limited number of total photons that can be collected (so-called photon budget). In addition, fast imaging is often needed to precisely map particle trajectories, in particular for those particles that move rapidly. High-speed imaging requires short exposure times that in turn give low SNR.

The different types of noise and the relationship between exposure time and SNR for cameras (e.g., charge-coupled device [CCD] and scientific complementary metal oxide semiconductor [sCMOS]) are described by the equation (Eq. 1; Fellers and Davidson, 2004) and the schematic in Box 1. Shot noise is due to the statistical nature of the rate of photons arriving on the camera, and the interval between arrival of photons is defined by Poisson statistics. This noise is captured by the factor  $(S + B)$  in the denominator of Eq. 1. An additional source of noise included in Eq. 1 is the background  $B$  generated by radiating fluorophores above and below the imaging plane, by scattered light in the microscope, as well as by autofluorescing objects in the sample. Dark noise is due to the statistical variation of the electrons thermally generated by the camera. For a given camera, we assume constant  $P$ ,  $QE$ ,  $N_r$ , and  $D_c$ . Dark noise is negligible for exposure times of 1 s or shorter compared with the typical read noise at standard frame rates of  $N_r = 1.4$  electrons root mean square (rms). Read noise is the noise generated by the amplifying electronics in a camera that converts the electrons generated at a pixel into a voltage output and then into a digital value via an analogue-to-digital converter. A key difference between CCD and sCMOS cameras is that the read noise is almost identical for each CCD pixel, whereas sCMOS pixels have read noise that is pixel-dependent. Note that read noise is squared in its contribution to the noise denominator in Eq. 1. Therefore only the signal and photon

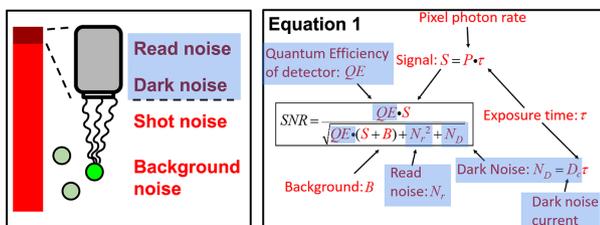
shot noise play significant roles in the variation of SNR with exposure time  $\tau$ .

Recognizing the various sources of noise and classifying their statistical features is at the core of several classic noise reduction algorithms. These include BM3D (Danielyan *et al.*, 2012) and nd-safir (Boulanger *et al.*, 2010). Both methods are based on collaborative filtering: they group together image patches with similar statistical properties and apply transformations that use all grouped patches to restore each grouped patch. The methods differ in how they define patch similarity, as well as in how they use the groups to restore patches. Nd-safir restores a patch by using the weighted average of the group, with more similar patches having greater weight in the average. BM3D weighs patch contributions within the group by its estimate of the noise they contain: less noisy patches carry greater weight. While effective in increasing the SNR of input images, these approaches are content agnostic: they use the statistical properties of the intensities within the input images, without regard to the image content. This makes them applicable to any data set as long as their assumptions are met. However, in some cases it may be beneficial to sacrifice generalizability and use approaches that are data set-specific to achieve a further increase in the SNR and superior restoration. Denoising approaches that make this trade-off are called content-aware: they collect and use information about the content of the data set and can achieve higher SNR and notable qualitative image improvements as a result of exploiting this additional information.

With the rapid improvement in graphical processing units (GPU), deep learning has revolutionized quantitative image analysis problems, including cell segmentation, object classification, and particle counting and tracking (Moen *et al.*, 2019). In addition, deep learning methods that use empirical knowledge on noise have a great potential to restore low-SNR data sets, as mentioned above. The content-aware image restoration (CARE) network developed by Weigert *et al.* (2018) and based on the U-net convolutional neural network (CNN) architecture (Ronneberger *et al.*, 2015) is a popular tool for image restoration. In its original implementation, CARE uses pairs of images with high and low noise levels to generate data-specific denoising networks. High-quality “ground truth” (GT) images with high SNR are not always available, and self-supervised approaches have been developed that eliminate the need for GT. One example is Noise2void (N2V) (Krull *et al.*, 2018, 2020). The innovation that allows N2V to denoise images with access to only a single noisy image (or set of images) is the introduction of blind-spot networks. The N2V network is given an image patch and tasked with predicting the center pixel’s value. A conventional U-net would learn to directly output the center pixel’s value, ultimately leaving the input image unchanged. In contrast, the N2V network has a blind spot at the center pixel, forcing it to infer the center pixel’s intensity from the surroundings. Krull *et al.* (2018, 2020) have shown that such blind-spot networks learn to remove pixel-wise independent noise.

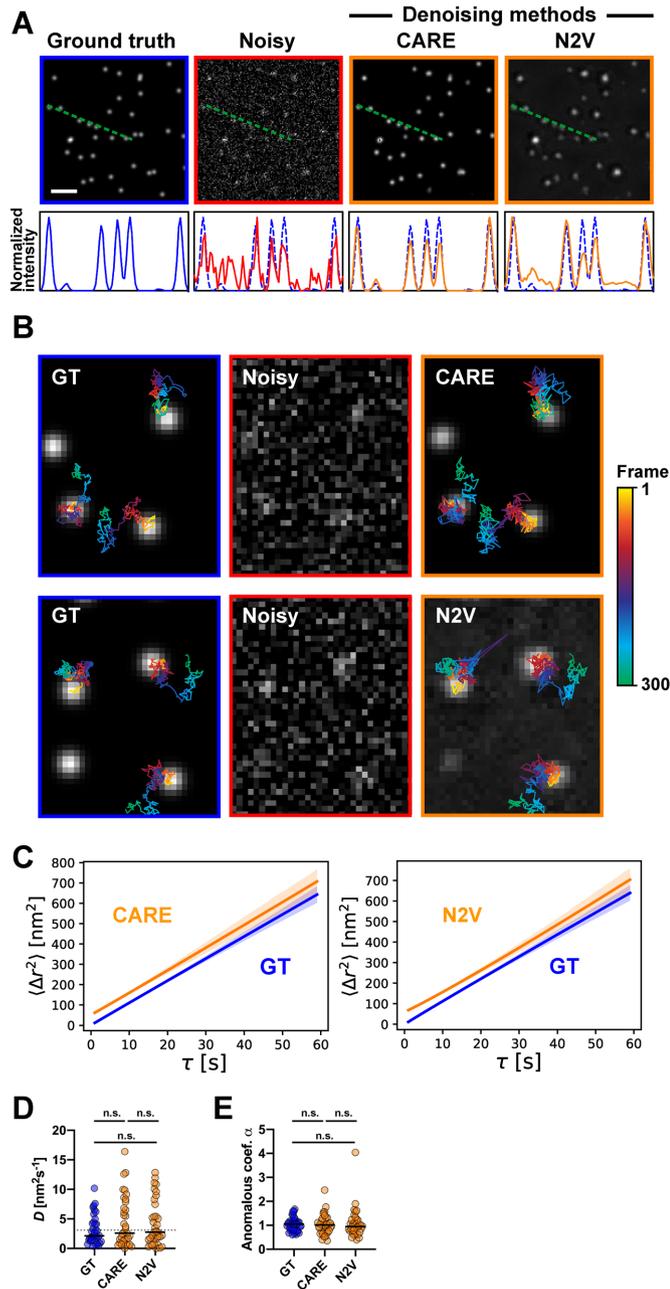
Image restoration with deep learning for quantitative microscopy is still in its infancy. While image improvement with CARE and N2V (among other deep learning approaches) is qualitatively impressive, to the best of our knowledge, the ability of these neural networks to restore data for quantitative analyses of molecular properties, such as diffusive behaviors, has not been rigorously evaluated. Particle tracking is an ideal test case to objectively assess the performance of image denoising approaches, specifically, their ability to restore meaningful biological information, beyond cosmetic improvements. Here, we compare tracking performances after

### Box 1:



Different sources of noise in (cellular) imaging. In the schematic, noise dependent on the camera is color-coded bordeaux on blue background, whereas noise dependent on the sample is coded in red. The bar indicates relative proportions of sample- and camera-related noise in typical imaging conditions. In Eq. 1,  $S = Pt$  is the signal ( $P$  is the incident photon flux [photons/pixel per second] and  $t$  is the exposure time [seconds]),  $QE$  represents the camera quantum efficiency (# electrons generated/incident photon),  $B$  is the background (same units as  $S$ ),  $D_c$  is the dark current value (electrons/pixel per second), and  $N_r$  represents camera read noise (electrons rms/pixel).

denoising with supervised and unsupervised CNNs, using synthetic image sequences of diffusing beads as well as chromatin dynamics as a cell-biological application.



**FIGURE 1:** Evaluation with synthetic data of the performance of supervised and unsupervised CNN denoising approaches for particle tracking. (A) Representative synthetic images of diffusing beads, without added noise (ground truth) or with noise added ( $\sigma = 0.5$ ; see *Materials and Methods*). Noisy images were denoised with different algorithms, as indicated. Intensity profiles along the dotted lines are shown. Scale bar, 2  $\mu\text{m}$ . (B) Illustration of particle trajectories in ground truth (GT) and denoised movies (CARE or N2V). Particles in noisy movies could not be tracked. (C) Averaged MSD curves from beads tracked in ground truth and denoised movies. (D) Diffusion values computed from the MSD curves. The dotted line indicates the theoretical value. (E) Anomalous diffusion coefficients ( $\alpha$ ) calculated from denoised and GT MSD curves. Statistical comparisons (D, E) using the Kruskal–Wallis test and Dunn’s multiple comparisons tests; n.s.,  $P > 0.99$ . Median values are indicated.

## RESULTS AND DISCUSSION

### Restoring images of diffusing particles: proof-of-concept with synthetic movies

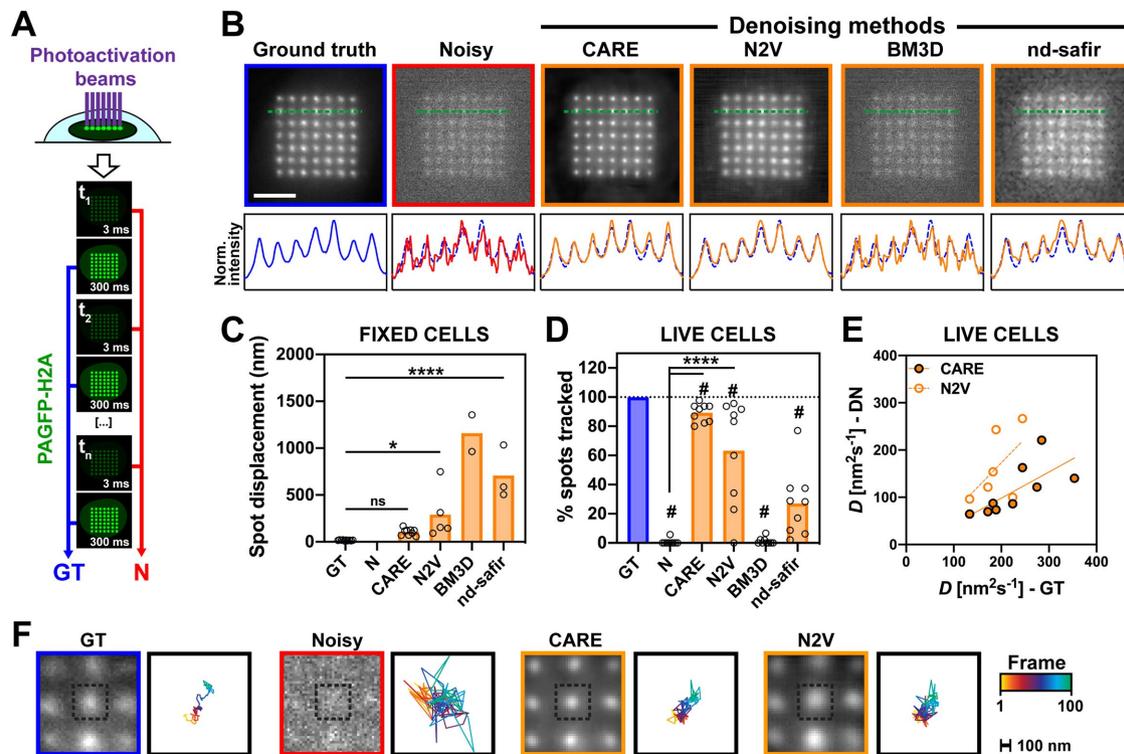
As an initial step to test whether CNN-based image denoising can improve accuracy in particle tracking experiments, we generated synthetic time lapses of moving beads. With this approach, the number of particles and their diffusion coefficient ( $D$ ), background shot noise, and readout noise were defined for each movie. In these artificial movies, all particles underwent Brownian motion and their mean squared displacement (MSD) was used to describe changes in the positions of the particles with respect to the characteristic time interval ( $\tau$ ). For denoising with CARE, pairs of movies were created that had either no added noise (ground truth) or different levels of added noise (Figure 1A and Supplemental Figure S1A). A subset of the data was used to train CARE networks, which were then applied to denoise the remaining noisy movies. N2V only requires a series of noisy images for training, without corresponding GT. The beads in both denoised and GT movies were detected and tracked using a single-particle tracking algorithm with cross-linking to develop the trajectories (see *Materials and Methods*).

As expected, both CARE and N2V improved noisy synthetic movies, revealing beads barely distinguishable by the human eye in images with high noise levels. This improvement can be appreciated with the intensity profiles shown in Figure 1A. Beads in movies with relatively low noise levels ( $\sigma = 0.2$ ) could be precisely tracked after denoising (Supplemental Figure S1, B and C). In movies with high noise levels ( $\sigma = 0.5$ ), for which tracking was not possible, CARE denoising enabled correct identification of  $\sim 75\%$  of the particles (Supplemental Figure S1D). Qualitatively, movies denoised with CARE showed individual particle trajectories closely resembling the ones extracted from the ground truth. Similarly, trajectories after N2V denoising were similar to those in the ground truth (Figure 1B). Quantitatively, MSD curves derived from movies processed with the CARE algorithm coincided with the ground truth (Figure 1C and Supplemental Figure S1C). Accordingly, diffusion coefficients and anomalous exponents derived from the denoised movies were consistent with the theoretically defined values ( $P > 0.2$ , one-sample Wilcoxon test), and not different from the ground truth (Figure 1, D and E). Similarly, untrackable movies could be analyzed after denoising with N2V. The resulting diffusion and anomalous exponent values were not different from the theoretical values or the values calculated from the ground truth movies. To further assess the accuracy of the denoised movies, we calculated the smallest distances between bead spot centers in GT and denoised movies. The cumulative values for these tracking errors are shown in Supplemental Figure S1F. After denoising with the CNNs (CARE and N2V), the mean cumulative localization error (over 300 frames) was  $\sim 50$  nm (or 1/5 of the size of the point spread function). It should be noted that this localization error did not deteriorate the restored particle trajectories; it just shifted the baseline of the MSD curves without affecting fitting results of  $D$  values (Michalet, 2010) (Figure 1C).

In summary, these experiments with synthetic images of diffusing particles show that CNN-based image restoration has the potential not only to improve the image quality, but also to recover the motion behavior of moving particles.

### Content-aware denoising for tracking chromatin microdomains in live cells

After establishing proof-of-concept with a synthetic data set, we applied the CARE and N2V softwares to noisy time lapses of slow-diffusing chromatin microdomains to evaluate whether and to what



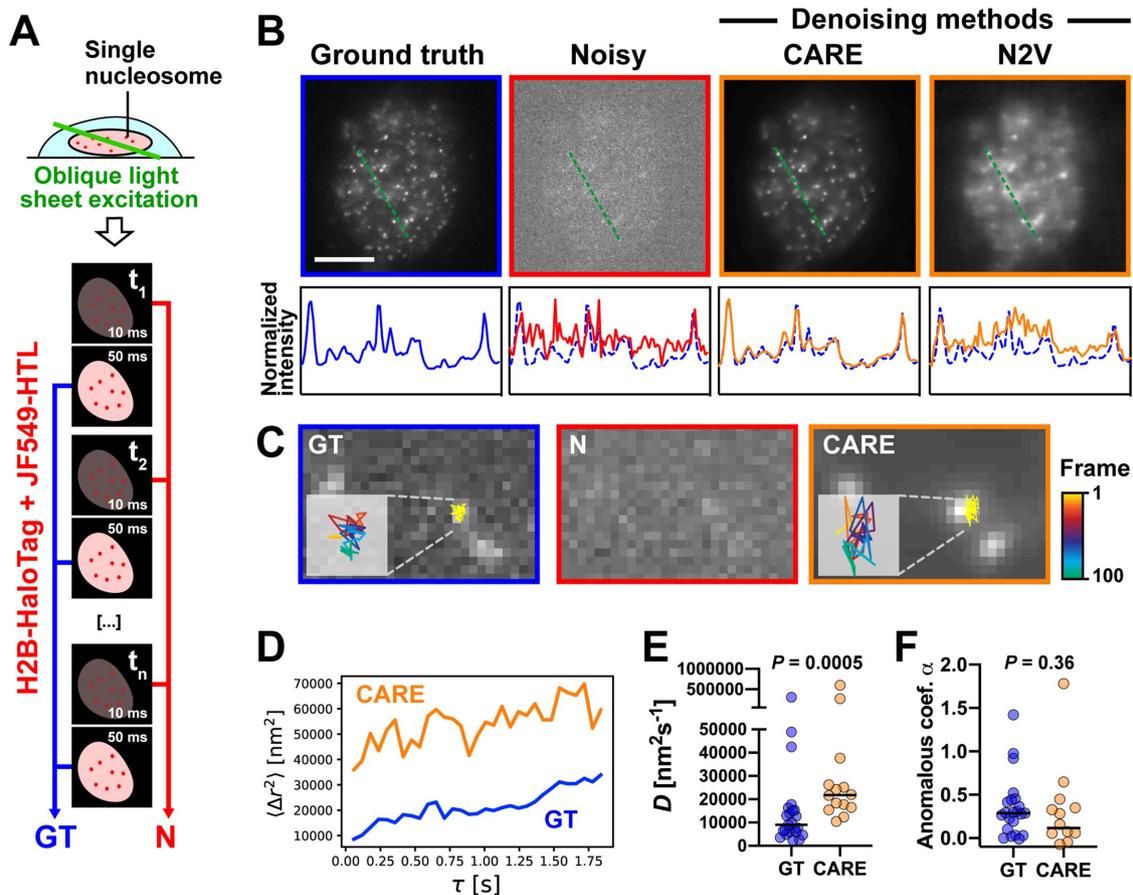
**FIGURE 2:** Image denoising to track chromatin microdomains in live-cell nuclei. (A) Schematic of the approach to photoactivate chromatin microdomains in U2OS cells expressing PAGFP-H2A. Rapid successions of short and long exposure times were collected. (B) Representative images of photoactivated chromatin microdomain lattices taken with a long exposure (ground truth), with a short exposure (noisy), and after denoising with different algorithms. Pixel intensity profiles are shown across a row of microdomains. Scale bar, 5  $\mu\text{m}$ . (C) Cumulative spot motions in fixed cells. No microdomain could be tracked in the noisy data set. (D) Percentages of photoactivated spots that could be tracked in noisy movies or after denoising with the different algorithms. For each cell, values are normalized to the number of spots tracked in the GT. (E) Chromatin diffusion ( $D$ ) in live cells, comparing GT and denoised (DN) movies. Each dot in the graphs (D, E) represents the average value for a cell. (F) Representative traces of microdomain trajectories. \*,  $P < 0.05$ ; \*\*\*\*,  $P < 0.0001$ ; ns, not significant (analysis of variance and Tukey's test). #,  $P < 0.05$  (one-sample  $t$  test; theoretical mean of 100). Mean values are indicated in the bar graphs.

extent CNN-based denoising can restore kinetic information in a biological data set. We used a custom optical setup based on diffractive optics to photoactivate  $7 \times 7$  lattices of chromatin microdomains in U2OS cells stably expressing histone H2A tagged with photoactivatable green fluorescent protein (PAGFP-H2A) (Bonin *et al.*, 2018) (Figure 2A). Interlaced movies were collected from the photoactivated lattices, alternating short and long exposure times. Because the time difference between adjacent frames (short long exposures) is small, we consider the image sequence taken with long exposures as the ground truth for the noisy (N) image sequence. We applied the same approach with different short exposures (10, 3, and 1 ms), corresponding to increasingly noisy images. To restore noisy image sequences, we used CARE networks trained on N/GT image pairs. Alternatively, we trained N2V networks tailored to individual noisy sequences with the goal to maximize the outcome of unsupervised denoising.

Qualitatively, images denoised (DN) with CARE or N2V were much sharper than the original noisy images and resembled the GT (Figure 2B and Supplemental Figure S2A). Accordingly, pixel intensity profiles across photoactivated spots revealed clear peaks that appeared to be similar in GT and DN images. Even for movies with the highest noise levels (1 ms exposure), improvement of the noisy images with CARE was impressive at first sight (Supplemental Figure 2B). Yet, closer examination revealed striking differences in the DN

outcomes, when comparing images restored from very low (1 ms) versus low (10 ms) exposure data. In DN images from very low exposures, fluctuations of background between frames were exacerbated, spot shapes appeared to have reduced complexity, and—most concerning—DN images occasionally included hallucinated spots inside and outside of the photoactivated lattice. Hence, denoising results need to be carefully evaluated in situations with extreme noise levels, in particular when periodic (and highly predictable) patterns are present in the images. Using patch sizes smaller than the periodicity of the repeated patterns is expected to mitigate these artifacts. However, this was not an option for our photoactivated grids with our current resolution. Increasing the proportion of images with missing pattern elements in the training set is another strategy that we will consider in future experiments. Finally, systemic parameter optimization (see below) with a reward function minimizing spot hallucination may further alleviate this problem.

To compare tracking accuracies achieved with GT, noisy, and DN data, image sequences were registered to cancel cell motions, and the center position of each chromatin microdomain in each image frame was determined by fitting with a two-dimensional (2D) Gaussian function. We used these positions to calculate MSD curves and diffusion coefficient  $D$  values (Bonin *et al.*, 2018). In fixed cells, we expect no microdomain motion; residual diffusion reflects drift of the microscope stage that was not properly



**FIGURE 3:** Restoration of the fast dynamics of nucleosomes captured with single-molecule imaging. (A) Schematic of the oblique light sheet imaging approach with U2OS cells expressing H2B-HaloTag. Rapid successions of short (10 ms) and longer (50 ms) exposure times were collected, leading to matching noisy (N) and ground truth (GT) movies. (B) Representative ground truth and noisy images, as well as images denoised with CNN algorithms. Pixel intensity profiles along the dashed lines are shown. Scale bar, 10  $\mu\text{m}$ . (C) Illustration of particle trajectories in GT and denoised (CARE) movies. Particles could not be tracked in noisy movies. (D) Representative MSD curves. (E) Diffusion values ( $D$ ) of single nucleosomes in GT and denoised movies. (F) Comparison of anomalous diffusion coefficients ( $\alpha$ ) derived from single-nucleosome MSD curves. Statistical comparisons (E, F) using Mann–Whitney test. Median values are indicated.

subtracted, combined with imprecise Gaussian fitting, leading to errors in spot center positions. As a first step to compare the DN methods, we calculated cumulative spot displacements in fixed-cell movies (Figure 2C and Supplemental Figure S2C). As expected (Bonin *et al.*, 2018), spots in GT movies from fixed cells barely moved ( $D = 2.2 \pm 1.3 \text{ nm}^2/\text{s}$ ). In contrast, the apparent spot motions in the corresponding noisy movies (10 ms exposure) were 10 times greater, and most spots from movies with greater noise levels (3 ms exposure) could not be tracked, partly due to failure in image registration (Supplemental Figure S2F). CARE and N2V denoising reduced the apparent spot motion of noisy images. Increased accuracy in spot center determination and improved image registration both contributed to this improvement in the DN movies (Supplemental Figure S2E). Next, we used the same approach with live cells. The proportion of chromatin microdomains that could be tracked significantly increased after denoising with CARE (Figure 2D and Supplemental Figure S2D). Denoising with N2V also increased the proportion of trackable spots, albeit to a lesser extent. To assess denoising outcomes,  $D$  values from DN movies were plotted against the corresponding values in GT (Figure 2E). Restoration of 3 ms movies to the 300 ms GT with CARE led to DN-derived diffusion values that were significantly correlated with

GT-derived  $D$  values ( $r = 0.72$ ,  $P = 0.029$ ). Moreover, individual microdomain traces from DN spots were qualitatively more similar to GT than to noisy source images (Figure 2F). Although CNN-based denoising (and CARE in particular) enabled tracking of >80% of the microdomains and yielded diffusion values correlated with those from GT, we note that  $D$  values in CARE-DN movies were systematically lower than their GT values. This effect may be due to the loss of dynamic changes in spot shape after denoising. Microdomain spots in our experiments have a typical size of 600 nm, and nonsymmetric shapes are usually observed in the GT, whereas images denoised with CARE display more symmetric spots with Gaussian profiles. In comparison, no such “smoothing” effect was observed with synthetic images (Figure 1A) or with single-molecule images (see next section; Figure 3B), in which spots are diffraction-limited with very nearly Gaussian profiles. It is likely that fluctuating structural variations are interpreted as noise by the software.

Another limitation in our comparisons is that the unsupervised CNN that we used (N2V) cannot identify or remove structural noise. To address this issue, we implemented Structured Noise2void (structN2V) (Broaddus *et al.*, 2020), a generalized version of N2V that uses blind-spot masks based on noise structure (Supplemental Figure S3). Autocorrelation analyses in image regions lacking cells

showed that, for the camera used to image chromatin microdomains, the main correlation was along a horizontal row of pixels (Supplemental Figure S3A). Indeed, horizontal stripes are clearly apparent in the background of images denoised with N2V (Figure 2B). Therefore, we used a horizontal blind mask to denoise chromatin microdomain movies with structN2V, which efficiently removed the horizontal stripe pattern, leaving only some vertical fiber-like features in the background of the images. Surprisingly, despite the qualitative improvement of structN2V (compared with N2V), tracking accuracy did not improve. On the contrary, apparent motions in fixed cells increased, fewer spots were tracked, and correlation of  $D$  with the ground truth decreased in movies denoised with structN2V compared with N2V (Supplemental Figure S3, B–D). Next, we used a combinatorial approach to select blind masks, using [–], [I], and [+]  
mask patterns with different pixel sizes (Supplemental Figure S3E). The background stripe artifacts completely disappeared when using plus-shaped [+]  
blind masks. Yet, tracking precision did not improve compared with N2V with any of the 15 blind masks tested (Supplemental Figure S3E). It is likely that fine-tuning the blind masks beyond our combinatorial approach (for example, using a Bayesian approach; see below) can improve restoration of particle dynamics. Still, our results highlight that nicer images are not necessarily better for quantitative microscopy, and in particular for particle tracking. Our interpretation is that increasing the size of the blind spot (N2V) to a larger mask (structN2V) leads to less information in the training process, causing a loss in accuracy for individual image features.

To benchmark the results obtained with CNN-based denoising, the same noisy image sequences were processed using two software solutions not based on deep learning, BM3D (Danielyan *et al.*, 2012) and nd-safir (Inria©) (Boulanger *et al.*, 2010). Both qualitatively and quantitatively, BM3D did not achieve CNN-based denoising outcomes (Figure 2, B–D, and Supplemental Figure S2A, C, and D). This was expected because this software is best suited for images with a large proportion of signal rather than dominance of background pixels. In addition, BM3D requires an estimate of the noise variance, which was not trivial to assess. Overall, results obtained with nd-safir with moderately noisy images (10 ms exposure) were comparable to those obtained with unsupervised deep learning (Supplemental Figure S2). With higher noise levels (3 ms exposure), movies denoised with nd-safir could not be tracked (Figure 2). The nd-safir software can handle 4D data sets. We may therefore have underestimated the performance of this software to restore tracking information because we did not exploit the temporal component of the movies. Nevertheless, the results suggest that CNN-based denoising outperforms classic approaches, at least for this specific quantitative microscopy application.

### Application of CARE for tracking single molecules in live cells

Next, we applied CNN-based denoising to single-molecule imaging by tracking the dynamics of individual nucleosomes in nuclei with stochastically labeled histone H2B. For these experiments, we used U2OS cells expressing H2B fused to the HaloTag, to which a fluorescent ligand can bind specifically (Liu *et al.*, 2018). Stochastic labeling of H2B was achieved by incubating U2OS H2B-HaloTag cells with a low concentration of fluorescent ligand, for a short amount of time. For imaging, live cells were illuminated by an oblique light sheet and the fluorescent signal was collected by a camera at a rate of up to 100 frames/s (Figure 3A). Compared with microdomains of chromatin, single nucleosomes have much faster kinetics that can be captured only by imaging at high frame rates (Nozaki *et al.*, 2017). And similarly to other SMT experiments, the

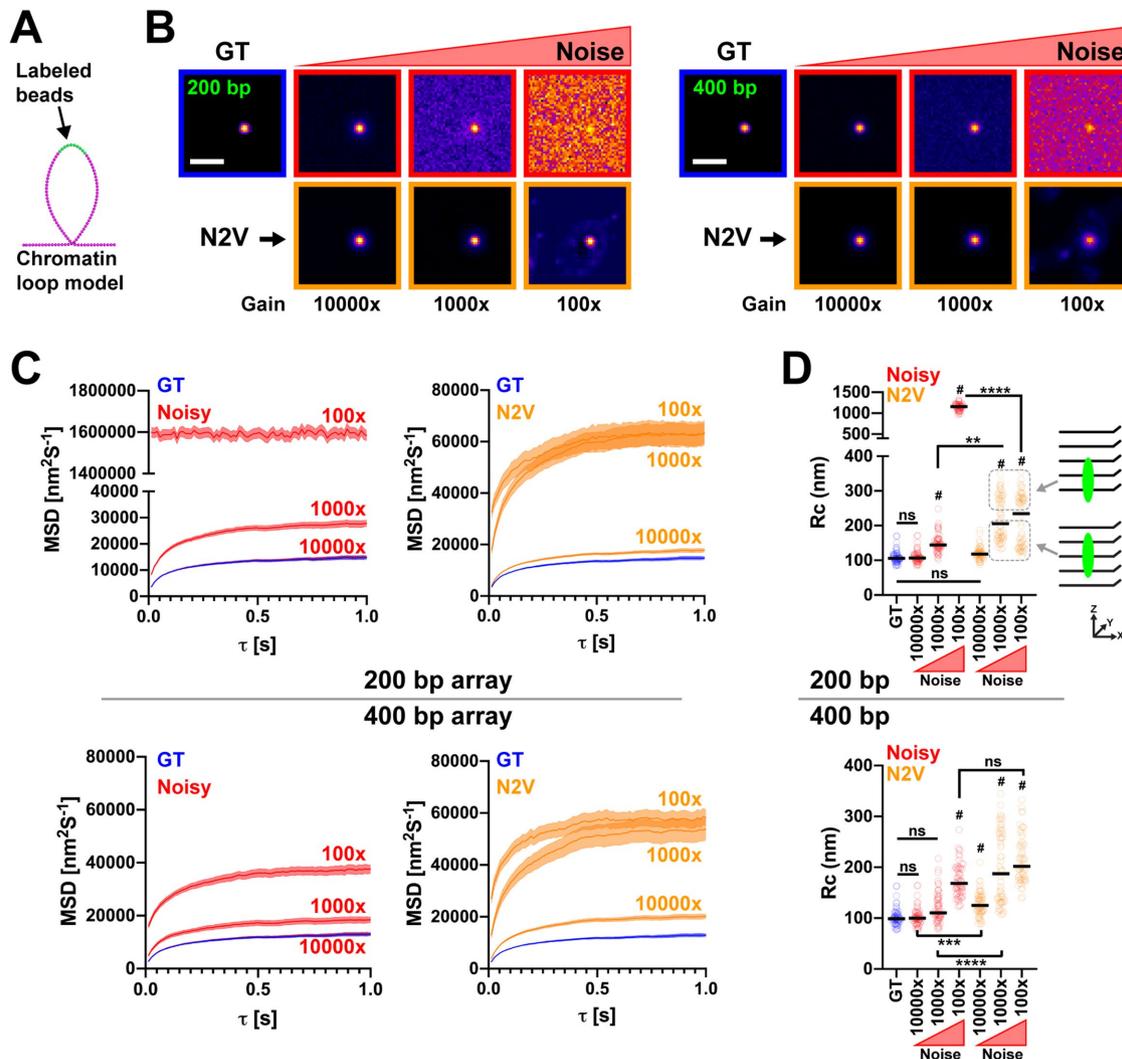
photon budget of each labeled nucleosome is limited, meaning that low illumination intensity is needed to capture time-lapse series of meaningful length. To train CNN denoising networks, we used fixed cells and captured matching pairs of low exposure (10 or 50 ms) and ground truth (1 s) images. For CARE, both noisy and ground truth images were used for supervised learning, while only the noisy data were used for N2V training. We used the same approach as the one used for chromatin microdomains, recording movies alternating short (10 ms) and longer (50 ms) exposures (Figure 3, A and B). The trained CARE and N2V networks were applied on the short exposure movies, while the longer exposure movies were used as ground truth.

Again, the performances of CARE exceeded that of N2V in restoring single-molecule images, as indicated by the intensity profiles of individual nucleosomes (Figure 3B). CARE restored both the localization and relative fluorescence intensity of nucleosome foci, whereas N2V partially recovered spots from the raw image, with intensity profiles indicating broadening of spot signals compared with the ground truth. CARE enabled tracking of ~59% of the nucleosome foci detected and tracked in ground truth images, whereas N2V restored about 15% of nucleosome foci in noisy movies. Therefore, for this data set and with the CNN training parameters that we used, the supervised denoising approach was better-suited than the unsupervised approach, as the pattern and statistical features of the noise are largely unknown. We found that minor changes to N2V training parameters have a strong impact on the denoising outcome. It is therefore likely that improvement of single-nucleosome tracking performance after N2V denoising is possible by further optimizing N2V training.

Next, we evaluated the dynamics of the nucleosomes following CNN-based image restoration. We had established that movies of nucleosomes taken with a 50 ms exposure can be used for single-molecule tracking with our system, whereas 10 ms movies could not be tracked due to the high background noise. First, we verified whether CARE denoising of “trackable” nucleosome movies induces artifacts. We obtained similar tracking results with 50 ms exposure images and the same images processed with the CARE network described above (Supplemental Figure S4). We then focused on high-noise movies. As illustrated in Figure 3C, trajectories of the same nucleosome tracked in both 50 ms exposure (GT) and denoised 10 ms exposure movies were qualitatively similar. More importantly, the calculated MSD curves based on the GT trajectories had slopes and shapes matching the ones derived from the 10 ms denoised trajectories (Figure 3D), with the y-axis shift due to the larger localization error of DN spots. Diffusion values derived from the denoised movies were about twofold larger than from the GT movie. This difference can be explained by the fact that motion blur is greater for longer exposure movies and leads to an underestimation of particle velocity, as observed previously by other groups (Amitai and Holzman, 2017; Miné-Hattab *et al.*, 2017; Shukron *et al.*, 2019). Statistically, the  $\alpha$  values were similar for the GT and DN movies (Figure 3F) and lower than 0.5, indicating that the subdiffusive behavior of nucleosomes (Nagashima *et al.*, 2019) was captured in DN movies.

### Denoising and tracking of 3D data set

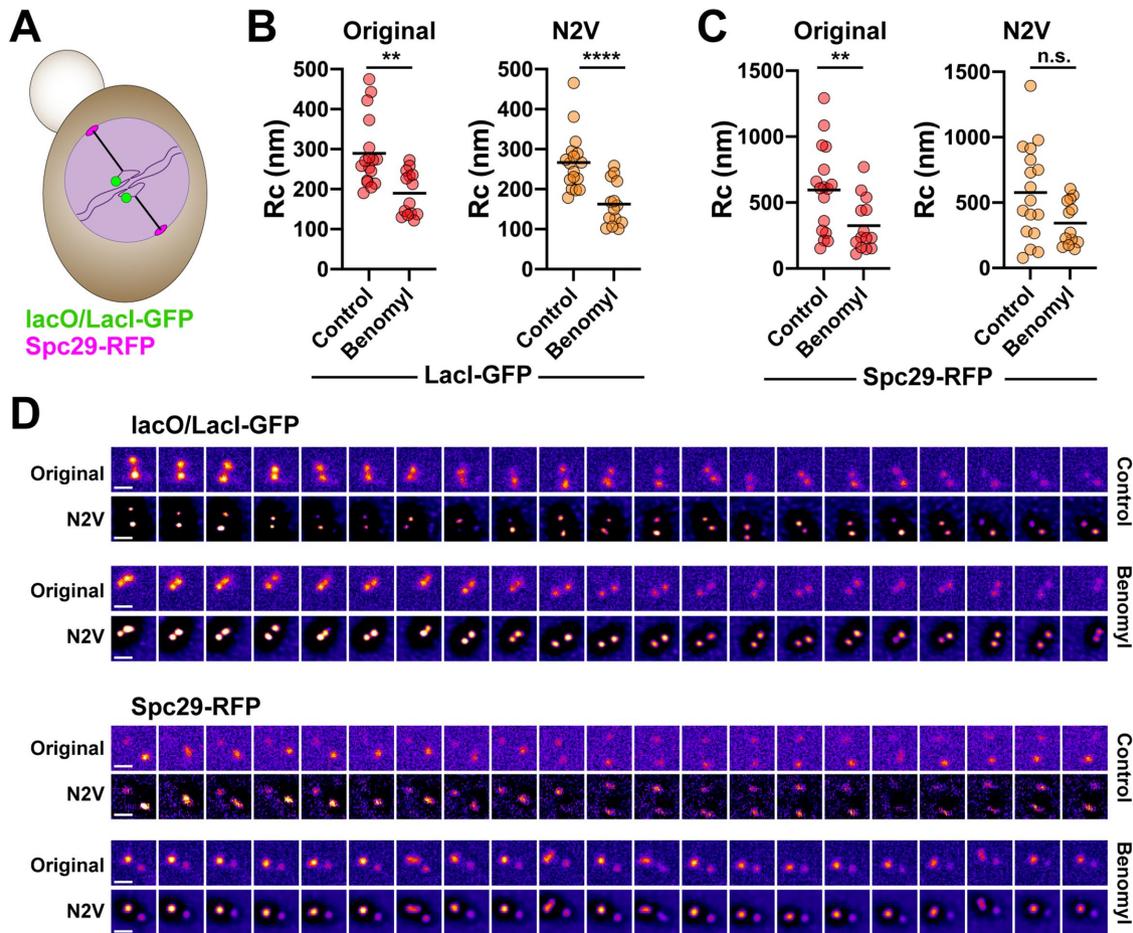
The results presented so far were all derived from simulated or experimental 2D images, with positions and trajectories projected and extracted from a single z-plane. While these types of measurements are practical and generally well-suited for flat samples, such as cell nuclei in monolayer cell cultures, 3D measurements improve particle tracking accuracy, in particular for round nuclei where the relationship between 2D and 3D distances deteriorates, and for short



**FIGURE 4:** Effect of CNN-based image volume restoration on tracking performance for single-chromosomal loci. (A) Schematic of the fluorescent labeling of a single loop from a ChromoShake simulation of the budding yeast pericentromere (Lawrimore *et al.*, 2016). (B) Representative images generated from ChromoShake simulations with 0 noise (GT) and normally distributed random noise with a SD of 5 AU. The noised images were generated with 100-, 1000-, and 10,000-fold gains to create images with different SNRs. Each image is displayed to show the image's full intensity range. The contrast differs between images. (C) Ensemble MSD curves of GT and denoised image foci of the simulated 200- (top) and 400 (bottom)-base-pair arrays. Each plot is composed of 64 different loops from a single-pericentromere simulation. Error bars represent SEM. Initial localization of the focus within the noised 100 $\times$  images of 200-base-pair array and of 400-base-pair array exceeded the cropping region of the 3D Gaussian fitting in 23.45% and 0.004% of images, respectively (see *Materials and Methods*). The large percentage of mislocalized foci in the 100 $\times$  images of the 200-base-pair array caused the large Rc values in D. Ground truth MSDs were calculated directly from simulation coordinates. (D) Comparison of radii of confinement (Rc) from GT and tracked images for 200- and 400-base-pair arrays. The schematic shows that the higher Rc values correspond to foci that are lower in the z-dimension for the denoised images of the 200-base-pair array. \*\*,  $P < 0.001$ ; \*\*\*,  $P < 0.0005$ ; \*\*\*\*,  $P < 0.0001$ ; #,  $P < 0.0001$  compared with GT; ns, not significant (Kruskal–Wallis and Dunn's multiple comparison test).  $N = 64$  simulated time lapses. Median values are indicated.

distances ( $<5 \mu\text{m}$ ) where the average 2D/3D discrepancy is  $\sim 30\%$  (Finn *et al.*, 2017). Tracking of genomic loci in the small, round, and tumbling nuclei of yeasts is one instance where 3D data sets are required (Gasser, 2002; Chac3n *et al.*, 2014; Lawrimore *et al.*, 2018). The radius of confinement (Rc), which corresponds to the area explored by a given genetic locus, is often used as a metric to quantify chromosome dynamics. As a gauge of the requirement for accuracy, the Rc of different genetic loci vary approximately threefold throughout an entire chromosome (Verdaasdonk *et al.*, 2013). Hence, Rc measurements need to be highly accurate to capture small differ-

ences in chromatin dynamics following a variety of perturbations. To assess particle tracking performance after CNN-based denoising of image volumes, we generated simulated 3D microscopy time lapses of the budding yeast pericentric region (Lawrimore *et al.*, 2016) with different noise levels (see *Materials and Methods*). These movies model DNA loops containing 200- or 400-base-pair fluorescent reporter operator arrays (Figure 4A). The loops exhibit confined motion due to the tethering of the simulated DNA at the centromere. The 400-base-pair-array signal is twice as bright as the 200-base-pair-array signal. Figure 4B shows representative images of the

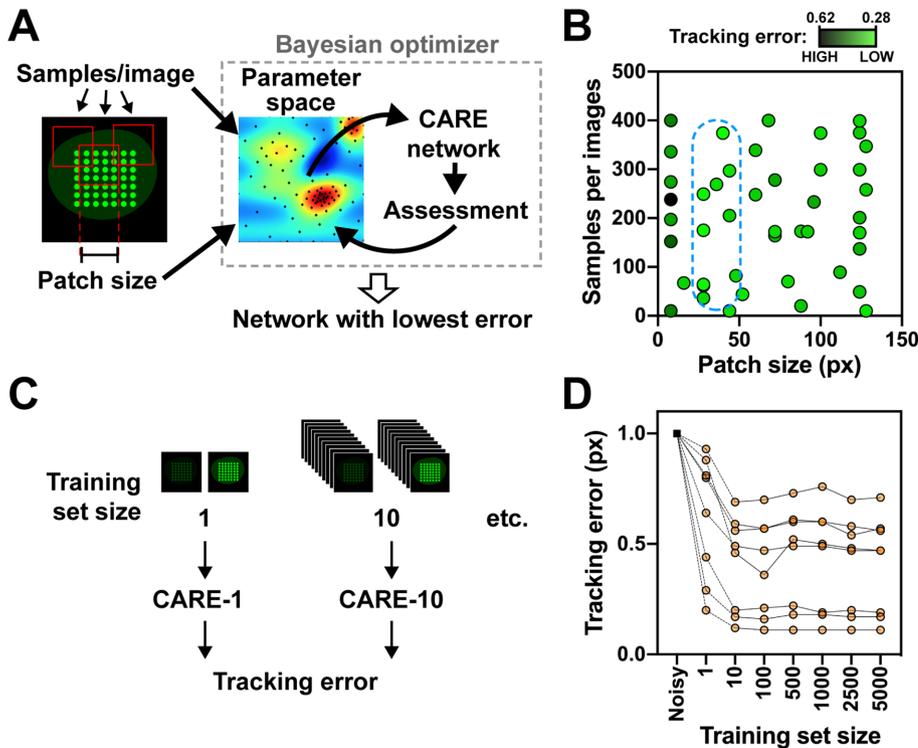


**FIGURE 5:** N2V denoising does not mask reduction of chromatin motion upon benomyl treatment. (A) Schematic of 10 kb lacO/LacI-GFP tandem repeat array located 1.8 kb from CEN15. The spindle body is labeled with Spc29-RFP. Z-stacks were collected, and maximal intensity projections were used for tracking the array and spindle body. (B, C) Radius of confinement (Rc) for sister LacI-GFP signals (B) and Spc29-RFP (C) for original images and images denoised with N2V. \*\*,  $P < 0.01$ ; \*\*\*\*,  $P < 0.0001$ ; ns, not significant (Mann-Whitney test). Mean values are shown. (D) Representative time-lapse montage of lacO/LacI-GFP sister foci signals and Spc29-RFP signals in original and N2V-denoised images of untreated or benomyl-treated, metaphase cells. Images are sum intensity projections with an interval of 30 s. Scale bars, 1  $\mu\text{m}$ .

simulated loci, with the different noise levels. Image sequences without any added noise are shown as ground truth comparison.

The content-aware denoising software that we have so far evaluated can also be trained on 3D data. Because ground truth images are rarely available for existing biological data sets, and considering that N2V is fully implemented in ImageJ, an image analysis software broadly used in the cell biology community, we focused on unsupervised denoising with N2V. The simulated hyperstacks were denoised using N2V, with one N2V network trained for each imaging condition on all images of that condition. The foci were tracked using a Gaussian fitting approach, and the MSDs at different time intervals were calculated for each of the tracks (Figure 4C). The plateau value of each MSD curve was converted into a Rc value. The ground truth MSDs and confinement radii were calculated directly from simulation coordinates and compared with values from noisy and N2V-denoised simulation movies. As expected, time lapses with low noise levels could be tracked precisely, yielding Rc values not statistically different from their respective ground truth values (Figure 4D and Supplemental Table S1). For the dimmest array (200 base pairs), processing with N2V did not significantly alter the tracking outcome (Figure 4, C

and D, compare 10,000 gain and GT). For the other brightness and noise conditions, denoising with N2V lead to significant overestimation of the Rc values, compared with both GT and noisy input movies. We think that this effect is due to increased localization uncertainty for individual beads after N2V denoising in this data set. Indeed, Rc values and localization errors are significantly correlated, for both the original and denoised images (Pearson correlation coefficient = 0.99 and 0.79, respectively; Supplemental Figure S5). Only in the most extreme condition (dimmer 200-base-pair array and highest noise level) did N2V denoising significantly improve tracking outcome. For these images, in which individual loci were barely distinguishable and tracking essentially failed (Figure 4D, top, 100 $\times$  gain), N2V denoising enabled tracking, yielding Rc values overestimated by a factor of two. The bimodal distributions present in the denoised 200-base-pair-array images are due to significantly more variation in the z-dimension tracking when the foci were near the bottom of the z-stack (see schematic in Figure 4D). We conclude that, when genomic loci can be tracked in 4D data sets, CNN-based denoising may introduce artifacts. Denoising becomes a more attractive solution when noise levels are very high (e.g., SNR < 1.5).



**FIGURE 6:** Optimization of CARE network parameters. (A) Schematic of the approach to simultaneously assess multiple parameters using Bayesian optimization. For chromatin microdomain tracking, the reward function minimized the cumulative tracking error. (B) Effect of the patch size and the number of training subsamples on tracking accuracy. The graph shows that the optimizer sparsely explored the entire parameter space before focusing on the most promising areas (dotted line). Each dot on the graph represents the performance (tracking accuracy) of a CARE network, mapping noisy (10 ms exposure) images to GT (300 ms) images from interlaced live-cell movies. (C) Approach to determine optimal size of the training set, with CARE CNNs trained on different numbers of image pairs. (D) The tracking error, expressed relative to the original noisy images, is shown as a function of the training set size. Each curve represents the data from one movie.

### Denoising single-chromatin loci in yeasts cells

We tested whether denoising using N2V would significantly alter the results of a single-particle tracking experiment. Time-lapse images of a 10 kb lacO/LacI-GFP array positioned 1.8 kb from centromere 15 and of spindle pole bodies labeled with Spc29-RFP (Figure 5A) were denoised using N2V. Low-dose benomyl treatment is known to restrict the motion of pericentric chromatin in yeast (Lawrimore *et al.*, 2015). The restriction of the motion of the foci in benomyl-treated cells was apparent in the original time-lapse images as well as in the denoised time lapses. The motion of the spindle pole bodies was not restricted upon benomyl treatment, and this result was recapitulated in the denoised time lapses (Figure 5B). While denoising did introduce tracking errors (outliers in Figure 5, B and C) for a few loci, these did not lead to statistically different results compared with the original images. Denoising maintained higher SNR over the time course (Figure 5D), meaning that implementing N2V could allow longer observations.

### A systematic approach to optimize the parameters of CNN networks

A nontrivial aspect of training CNN networks such as CARE is choosing the network's parameters and deciding on the size of the training set. While some hyperparameters such as the learn-

ing rate may be application-invariant and thus are generally kept as described by the authors (Weigert *et al.*, 2018), other network architecture parameters such as the restoration patch size are more application-specific. Refining these parameters, which are accessible and implementable by biological users, with a system may improve the restoration outcomes. Yet relying on trial-and-error is time consuming. To rigorously identify parameters optimal for our type of data, we employed Bayesian optimization (Nogueira, 2014) (Figure 6A). In this approach, the network's parameter space is modeled using Gaussian processes. This optimization can find parameter values that lead to the best performance at the task at hand while minimizing the number of evaluations. The approach is thus especially useful when the evaluation of the target function is expensive, as in our case in which an image restoration network needs to be trained and its predictions tracked at each evaluation step of the optimizer. We used three different objective functions for assessment: cumulative tracked motion in fixed cells, cumulative tracking error in live cells, and relative number of spots tracked. Figure 6B shows the sets of optimal parameter values found by the Bayesian optimization process identified using these target functions. Compared with the default parameter set of the CARE implementation used, the parameters found through this optimization resulted in a model producing restorations with a 12.8% lower cumulative tracking error.

Producing large training sets for the supervised deep learning approach with matching noisy and clean images can be time consuming. To investigate the effect of training set size on the restoration outcome, we trained CARE networks with identical hyperparameters on differently sized training sets and quantified the change in tracking error (Figure 6, C and D). As expected, we found that models trained with more training data had a lower cumulative tracking error. A performance plateau was achieved with a surprisingly small number of training image pairs (10–100). This suggests that, at least in our experimental conditions, significant improvement in image quality and tracking precision can be achieved with a small data set. We do not exclude the possibility that further improvements would have been achieved with a much larger training set but considered that a very large set of matched noisy/GT image pairs is experimentally and computationally impractical.

### Conclusion

We evaluated the performance of content-aware deep learning methods for denoising microscopy image sequences, using particle tracking outcomes as an objective assessment. In contrast to conventional denoising approaches, CNN-based image restoration makes little or no assumptions on noise. No preacquisition of camera-based noise features or calibrations is needed, which facilitates the implementation of these methods. By learning the noise pattern

of a given image (or image pairs), both N2V and CARE remove noise regardless of its source (shot noise, dark noise, readout noise, etc.), with the notable exception of structural noise for N2V. Overall, we find that the deep-learning methods can restore biophysical information from very noisy data sets, highlighting their potential for quantitative microscopy. As expected, the supervised method (CARE) generally performed better than the unsupervised methods (N2V and structN2V). Yet, results with N2V and stN2V are promising because an unsupervised learning approach is the only option when paired training data sets are not obtainable. We also do not exclude that better N2V performances may be achieved with longer training times (which were not practically achievable for this study). Our results are very encouraging because individual chromophores in single-molecule tracking experiments have a limited photon budget, necessitating minimal excitation light during image acquisition. CNN-based denoising in our analyses was particularly efficacious to restore single-image planes, with mixed results obtained for image volumes. We also conclude that these methods are mostly relevant when noise levels are high and note the important caveat that, in contrast to classic denoising methods, content-aware deep-learning approaches can fabricate biologically irrelevant information that needs to be carefully evaluated. We find that image registration often fails when SNR is low, which is a major issue for particle tracking. In contrast, spot identification and Gaussian fitting are surprisingly resilient (except in cases of extremely high noise). Hence, it will be interesting to test a mixed approach, where content-aware denoising is used to define registration parameters; registration and particle identification and tracking would be done on the original image sequence. We anticipate further improvements in deep-learning denoising methods as the field rapidly expands and propose that the approach presented here will be useful to rigorously assess their performances.

## MATERIALS AND METHODS

### Synthetic data to evaluate denoising performances for particle tracking in 2D

A synthetic model was designed by our lab where the number of particles, diffusion coefficient, and noise can be defined for each movie. MSD was used to characterize particle motions. MSD is defined by the following equations:

$$\text{MSD}(\tau) = \langle \Delta r(\tau)^2 \rangle = \langle [r(t+\tau) - r(t)]^2 \rangle \quad (2)$$

In two dimensions,

$$\text{MSD}(\tau) = 4D\tau^\alpha \quad (3)$$

The term  $D$  in Eq. 3 is the diffusion coefficient of the particle, and the exponent  $\alpha$  is a unitless parameter that characterizes the type of diffusion;  $\alpha = 1$  for simple diffusion, and  $\alpha = 1/2$  for a stretch of beads in a long-chain polymer (Osmanović and Rabin, 2017). In a Brownian motion model, the MSD is dependent on the size of the moving object as well as the mechanical and physical properties of the medium, as described in Eq. 4 (the Stokes–Einstein equation):

$$D = kT/6\pi\eta R \quad (4)$$

Here  $\eta$  is the viscosity of the medium,  $T$  is the temperature,  $R$  is the particle radius, and  $k$  is the Boltzmann constant. To match experimental observations, the pixel size of the synthetic data was set as 80 nm. The particle was simulated as a 2D Gaussian function with a radius of 60 nm, and the intensity level was scaled from 0 to 1 with a Poisson distribution. To avoid merging of multiple foci, the  $D$  value

was set to 3.1 nm<sup>2</sup>/s. Movies including 49 particles lasting for 300 frames were simulated with a temporal interval of 1 s. The noise level of synthetic images was simulated as Gaussian noise, where the standard derivation ( $\sigma$ ) varies as  $\sigma = 0, 0.1, 0.2, 0.3, 0.4, 0.5,$  and  $0.6$ . Synthetic movies without noise ( $\sigma = 0$ ) were used as ground truth, and the coordinates of each particle were also used as ground truth for MSD curves. For restored images, the motion of the particle was tracked by a home-written algorithm (details below). A cut-off distance of 150 nm was used to define matching particles in two corresponding movies.

### Image generation and tracking of simulated genomic loci

We used ChromoShake (Lawrimore *et al.*, 2016) simulations of the budding yeast pericentric region to simulate yeast genomic loci. These simulations were converted into synthetic images using Microscope Simulator 2 (Quammen *et al.*, 2008). In Microscope Simulator 2, a custom point spread function was generated using the “Calculated Gibson–Lanni Widefield PSF” method for a 100 $\times$  magnification, 1.49 numerical aperture objective. The custom point spread function was convolved with either 10 or 20 consecutive monomer units, representing 200- or 400-base-pair fluorescent reporter operator arrays, respectively. These were positioned at the apex of 64 different loops in the pericentric simulations by first removing the header from the ChromoShake outfile and converting the coordinates from meters to microns. The resulting text file was converted to a series of Microscope Simulator 2 model files. The model files were converted to TIFF stacks. The simulated TIFF stacks were concatenated into 4D hyperstacks (50  $\times$  50 pixels, 5 z-planes, 200 nm z-step, 201 time points) using FIJI (Schindelin *et al.*, 2012). All simulations were converted into images containing either no noise (ground truth), or randomly distributed noise with a SD of 5 AU. The gain of the signal of the noisy images was set to 100-fold, 1000-fold, and 10,000-fold, to generate images with different SNRs.

The simulated DNA loci were tracked in each 4D hyperstack using a custom MATLAB code that locates the brightest voxel in a z-stack, crops a 15  $\times$  15  $\times$  5 region surrounding the brightest voxel, and uses MATLAB’s `lsqcurvefit` function to fit a 3D Gaussian function to the cropped region. The center of the fitted Gaussian function was calculated for each time point to create a single track per time lapse. The ground truth track for each time lapse was calculated by taking the mean position of all the labeled masses directly from the simulation model XML file using a custom MATLAB program. The MSD and radius of confinement were calculated by custom MATLAB programs.

### Mammalian cell culture

U2OS osteosarcoma cells were cultured in DMEM supplemented with 10% fetal bovine serum (Sigma) at 37°C, 5% CO<sub>2</sub>. Cells were seeded in 35 mm glass-bottom dishes (MatTek) at 100,000 cells per dish and imaged 48 h after seeding. U2OS cells stably expressing PAGFP-H2A (Bonin *et al.*, 2018) were used to track chromatin microdomains. For single-nucleosome imaging, we generated U2OS cells stably expressing H2B fused to the HaloTag by transfection of the pBREBAC-H2BHalo plasmid (Addgene plasmid #91564) using Lipofectamine 3000 (ThermoFisher) followed by clonal selection with geneticin. Before live-cell imaging, H2B-HaloTag U2OS cells were incubated with 10 pM fluorescent JF 459 HaloTag ligand (Grimm *et al.*, 2015) for 1 h, washed three times with phosphate-buffered saline, and incubated in DMEM without phenol red for at least 30 min. This concentration of dye proved to be optimal for imaging and tracking. For fixed imaging, cells were imaged after fixation with Formalin (Sigma #HT5011; 20 min).

## Chromatin microdomain tracking

Grids of photoactivated chromatin microdomains ( $7 \times 7$ ) were generated in U2OS PAGFP-H2A cells with a custom diffractive optical element module (Bonin *et al.*, 2018) inserted into the condenser arm of an inverted Olympus IX83 microscope. Cells were kept at 37°C in the custom-built enclosure of the microscope. Images were taken with a 60× oil lens (N.A. = 1.35) and an sCMOS camera (ORCA-Flash 4.0 v3; Hamamatsu Photonics; dark current  $D_c = 0.2$  electrons/pixel per second when cooled to  $-15^\circ\text{C}$ ), using the CellSens software. Images were registered using the StackReg plugin in ImageJ (Thévenaz *et al.*, 1998). Tracking of chromatin microdomains was done in MATLAB, as described (Bonin *et al.*, 2018).

## Single-nucleosome tracking

Motion of single nucleosomes in live U2OS cells was tracked using a custom-built oblique light sheet microscope, based on ASI's rapid automated modular microscope (Applied Scientific Instrumentation). A solid-state laser (559 nm, 30 mW; Olympus) was focused by a tube lens to the side of the back pupil of an oil immersion objective (60× N.A. = 1.2; Olympus), leading to a thin sheet of light that illuminated the cell. Fluorescent signals were collected by the same objective and further filtered by a multiband emission filter (69013M; Chroma). Finally, the fluorescent signal was detected by an ORCA-Flash 4.0 sCMOS camera. Cells were maintained in a physiological environment using a live-cell imaging chamber (INU-TIZ-F1; Tokai Hit). The microscope system and the time-course image acquisition were controlled by the open-source software MicroManager. Single-nucleosome motions were analyzed with a custom single-molecule image analysis platform, smCellQuantifier. Specifically, the localization of each nucleosome locus was detected and fitted with a 2D Gaussian function. Trajectories of single nucleosomes were established with a multitemporal association tracking algorithm (Shafique and Shah, 2005; Winter *et al.*, 2012). The MSD of each focus was calculated based on the trajectory profile, and the diffusion coefficient  $D$  as well as the anomalous coefficient  $\alpha$  were calculated with Eq. 3.

## Tracking genomic loci in yeast

Budding yeast strain KBY8065 (Mat a CEN15(1.8)-GFP[10kb] ade2-1, his3-11, trp1-1, ura3-1, leu2-3,112, can1-100, LacINLSGFP:HIS3, lacO::URA3, Spc29RFP:Hyg) was grown in liquid yeast extract peptone dextrose at 24°C. Cells were imaged in liquid yeast complete medium at 24°C. Time-lapse images were acquired on an Eclipse Ti wide-field inverted microscope (Nikon) with a 100× Apo TIRF 1.49 NA objective (Nikon) and a Clara CCD camera (Andor) using the Nikon NIS Elements imaging software. Time lapses were 10 min in duration with 30 s intervals. At each interval, a seven-step Z-stack of 400-nm step size was acquired in the GFP, RFP, and Trans channels.

Metaphase yeast cells (medium budded cells with two Spc29-RFP foci) were cropped by hand from the original time-lapse images. Both original and denoised time lapses were automatically tracked using a custom MATLAB program. The motion of the two sister lacO/LacI-GFP foci and the two sister Spc29-RFP foci were the motion of one focus relative to the other as in Chacón *et al.* (2014), and the radius of confinement was calculated by a custom MATLAB program. For the images shown as illustrations, the heterogeneous background was subtracted with the rolling-ball method (10 pixel radius) in FIJI. Tracking results are from original and N2V-denoised images.

## CNN training and Bayesian optimization of network parameters

For synthetic bead data (2D), 100 pairs of noisy and GT images were used for training the CARE algorithm, using a patch size of

$128 \times 128$  pixels. N2V networks were trained on images with high noise levels ( $\sigma = 0.5$ ), using a patch size of  $64 \times 64$  pixels, 100 epochs, and 100 steps per epoch.

For restoration of chromatin microdomain images, dedicated CARE networks were trained for each exposure time condition (1, 3, 10 ms) using pairs of cropped images taken from fixed cells. These training pairs were obtained by alternately imaging at the target exposure time (e.g., 3 ms) and an exposure time sufficient to achieve a high SNR (e.g., 300 ms) for 100 times and subsequently cropping image stacks centered at the grid of photoactivated chromatin microdomains. The parameters for patch size ( $28 \times 28$  pixels) and samples per image (64) were chosen by applying the Bayesian optimization implementation by Nogueira (2014). This method initially randomly samples the hypothesis space and then fits Gaussian processes to the observations. An acquisition function then determines the next point in the parameter space that would improve this model of the parameter space the most. Using this active learning approach, near-optimal parameter values are found without having to exhaustively search the parameter space. In our application, the parameter space is 2D with the number of samples taken per image as one, and the side length of the samples as the other dimension. The reward function of the optimizer is to find the set of parameters that minimize cumulative tracking error. We calculated tracking error as pixels per microdomain per frame: we summed the magnitude of the difference between the motion vector tracked in the ground truth and the motion vector tracked in the denoised image for each frame delta across all spots and then divided by the number of frame deltas and spots.

N2V networks for use with the chromatin microdomain images were trained on the cropped target image stacks directly, with one network trained for each image stack. We evaluated performance on training N2V networks on the full images ( $2048 \times 2048$  pixels) but found no improvement. Rather, training on the full images significantly increased training time. For denoising with structN2V, we first assessed the spatial autocorrelation of the noise using the MATLAB autocorr2d function and single images that were away from any cells. For both CARE and N2V we used a 90%–10% train-validation split.

To restore single-nucleosome images, a CARE network was trained using 100 pairs of fixed-cell images, captured at 1 s and 10 ms. N2V was trained using stacks of fixed images (100; 10 ms exposure). We optimized patch size ( $128 \times 128$  pixels for CARE and  $64 \times 64$  for N2V) before training the networks. Because we had high-quality fixed-cell images (1 s exposure) as the GT, we were able to decide on the most accurate network.

For yeast genomic loci and spindle pole bodies, images were processed using 3D N2V to generate denoised time lapses. For both simulated genomic loci and spindle pole body time lapses, one N2V network was trained for each observation, with a patch size of 32, 32, and 4 pixels for X, Y, and Z, respectively.

## Statistical analyses

Statistical analyses were done using GraphPad Prism 8. The D'Agostino & Pearson omnibus normality test was used to test for normality. Nonparametric tests were used if the data did not pass the normality test (at  $\alpha = 0.05$ ). Statistical tests are indicated in the figure legends.  $P$  values  $\leq 0.05$  were considered significant. All statistical tests were two sided.

## ACKNOWLEDGMENTS

We thank Jacques Neefjes (the Netherland Cancer Institute) for providing the PAGFP-H2A DNA construct and Zhe Liu (Janelia Research Campus, Howard Hughes Medical Institute) for sharing the LZ10 PBREBAC-H2B Halo DNA plasmid. We are grateful to George

Holzwarth for his comments on the manuscripts, for his contribution in developing MATLAB code, and for productive discussions. We also acknowledge Clayton Seitz and Hua Lin for contributing the single-molecule tracking software and the synthetic image generator for this project. This work was funded by the National Cancer Institute (U01CA214282 and P30CA012197 to the Wake Forest Baptist Comprehensive Cancer Center).

## REFERENCES

- Amitai A, Holcman D (2017). Polymer physics of nuclear organization and function. *Phys Rep* 678, 1–83.
- Bonin K, Smelser A, Moreno NS, Holzwarth G, Wang K, Levy P, Vidi P-A (2018). Structured illumination to spatially map chromatin motions. *J Biomed Opt* 23, 1–8.
- Boulanger J, Kervrann C, Bouthemy P, Elbau P, Sibarita J-B, Salamero J (2010). Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans Med Imaging* 29, 442–454.
- Broadus C, Krull A, Weigert M, Schmidt U, Myers G (2020). Removing structured noise with self-supervised blind-spot networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA: IEEE, 159–163.
- Carlton PM, Boulanger J, Kervrann C, Sibarita JB, Salamero J, Gordon-Messer S, Bressan D, Haber JE, Haase S, Shao L, et al. (2010). Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc Natl Acad Sci USA* 107, 16016–16022.
- Chacón JM, Mukherjee S, Schuster BM, Clarke DJ, Gardner MK (2014). Pericentromere tension is self-regulated by spindle structure in metaphase. *J Cell Biol* 205, 313–324.
- Chenouard N, Smal I, de Chaumont F, Maška M, Sbalzarini IF, Gong Y, Cardinale J, Carthel C, Coraluppi S, Winter M, et al. (2014). Objective comparison of particle tracking methods. *Nat Methods* 11, 281–289.
- Danielyan A, Katkovnik V, Egiazarian K (2012). BM3D frames and variational image deblurring. *IEEE Trans Image Process* 21, 1715–1728.
- Fellers TJ, Davidson MW (2004). CCD noise sources and signal-to-noise ratio. optical microscopy primer, Florida State University, Molecular Expressions, <http://microMagnetFsuEdu/primer>.
- Finn EH, Pegoraro G, Shachar S, Misteli T (2017). Comparative analysis of 2D and 3D distance measurements to study spatial genome organization. *Methods* 123, 47–55.
- Gasser SM (2002). Visualizing chromatin dynamics in interphase nuclei. *Science* 296, 1412–1416.
- Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, Patel R, Macklin JJ, Normanno D, Singer RH, et al. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* 12, 244–250.
- Krull A, Buchholz TO, Jug F (2018). Noise2Void - Learning Denoising From Single Noisy Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2129–2137.
- Krull A, Vičar T, Prakash M, Lalit M, Jug F (2020). Probabilistic Noise2Void: unsupervised content-aware denoising. *Front Comput Sci* 2, 5.
- Lawrimore J, Aicher JK, Hahn P, Fulp A, Kompa B, Vicci L, Falvo M, Taylor RM 2nd, Bloom K (2016). ChromoShake: a chromosome dynamics simulator reveals that chromatin loops stiffen centromeric chromatin. *Mol Biol Cell* 27, 153–166.
- Lawrimore J, Doshi A, Friedman B, Yeh E, Bloom K (2018). Geometric partitioning of cohesin and condensin is a consequence of chromatin loops. *Mol Biol Cell* 29, 2737–2750.
- Lawrimore J, Vasquez PA, Falvo MR, Taylor RM 2nd, Vicci L, Yeh E, Forest MG, Bloom K (2015). DNA loops generate intracentromere tension in mitosis. *J Cell Biol* 210, 553–564.
- Liu H, Dong P, Ioannou MS, Li L, Shea J, Pasolli HA, Grimm JB, Rivlin PK, Lavis LD, Koyama M, et al. (2018). Visualizing long-term single-molecule dynamics in vivo by stochastic protein labeling. *Proc Natl Acad Sci USA* 115, 343–348.
- Magidson V, Khodjakov A (2013). Circumventing photodamage in live-cell microscopy. *Methods Cell Biol* 114, 545–560.
- Michalet X (2010). Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys Rev E Stat Nonlin Soft Matter Phys* 82, 041914.
- Miné-Hattab J, Recamier V, Izeddin I, Rothstein R, Darzacq X (2017). Multi-scale tracking reveals scale-dependent chromatin dynamics after DNA damage. *Mol Biol Cell* 28, 3323–3332.
- Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D (2019). Deep learning for cellular image analysis. *Nat Methods* 16, 1233–1246.
- Nagashima R, Hibino K, Ashwin SS, Babokhov M, Fujishiro S, Imai R, Nozaki T, Tamura S, Tani T, Kimura H, et al. (2019). Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *J Cell Biol* 218, 1511–1530.
- Nogueira F (2014). Bayesian optimization: open source constrained global optimization tool for Python, <https://github.com/fmfn/BayesianOptimization>.
- Nozaki T, Imai R, Tanbo M, Nagashima R, Tamura S, Tani T, Joti Y, Tomita M, Hibino K, Kanemaki MK, et al. (2017). Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol Cell* 67, 282–293.e7.
- Osmanović D, Rabin Y (2017). Dynamics of active Rouse chains. *Soft Matter* 13, 963–968.
- Quammen CW, Richardson AC, Haase J, Harrison BD, Taylor RM 2nd, Bloom KS (2008). FluoroSim: A Visual Problem-Solving Environment for Fluorescence Microscopy, Eurographics Workshop Vis Comput Biomed 2008, 151–158.
- Ronneberger O, Fischer P, Brox T (2015). U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, ed. N Navab et al., Cham, Switzerland: Springer International Publishing, 234–241.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682.
- Shafique K, Shah M (2005). A noniterative greedy algorithm for multi-frame point correspondence. *IEEE Trans Pattern Anal Mach Intell* 27, 51–65.
- Shukron O, Seeber A, Amitai A, Holcman D (2019). Advances using single-particle trajectories to reconstruct chromatin organization and dynamics. *Trends Genet* 35, 685–705.
- Thévenaz P, Ruttimann UE, Unser M (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* 7, 27–41.
- Verdaasdonk JS, Vasquez PA, Barry RM, Barry T, Goodwin S, Forest MG, Bloom K (2013). Centromere tethering confines chromosome domains. *Mol Cell* 52, 819–831.
- Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broadus C, Culley S, et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* 15, 1090–1097.
- Winter MR, Fang C, Banker G, Roysam B, Cohen AR (2012). Axonal transport analysis using Multitemporal Association Tracking. *Int J Comput Biol Drug Des* 5, 35–48.