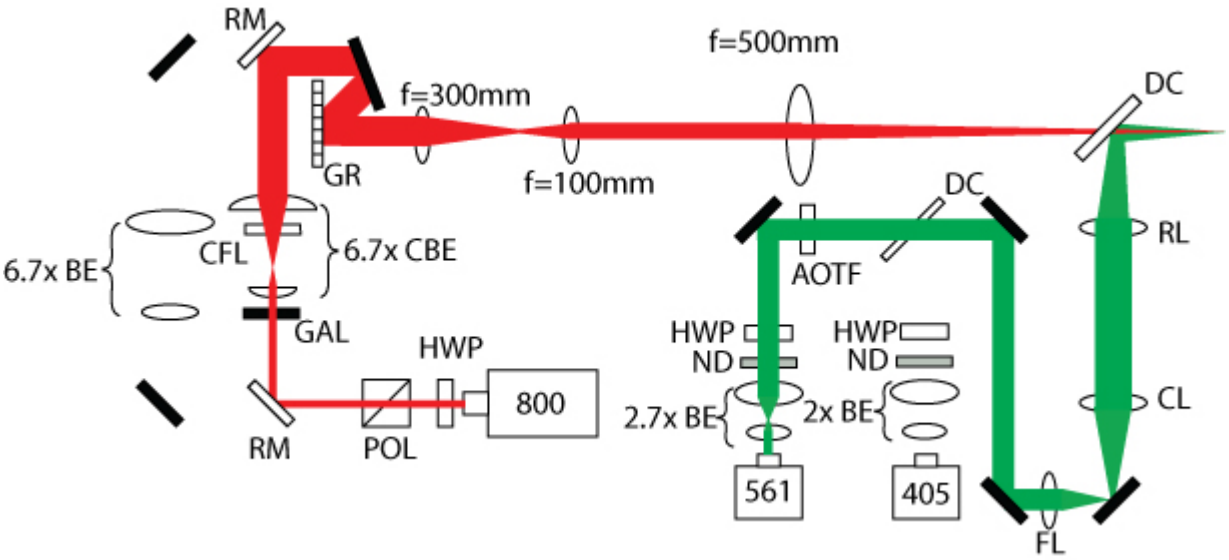
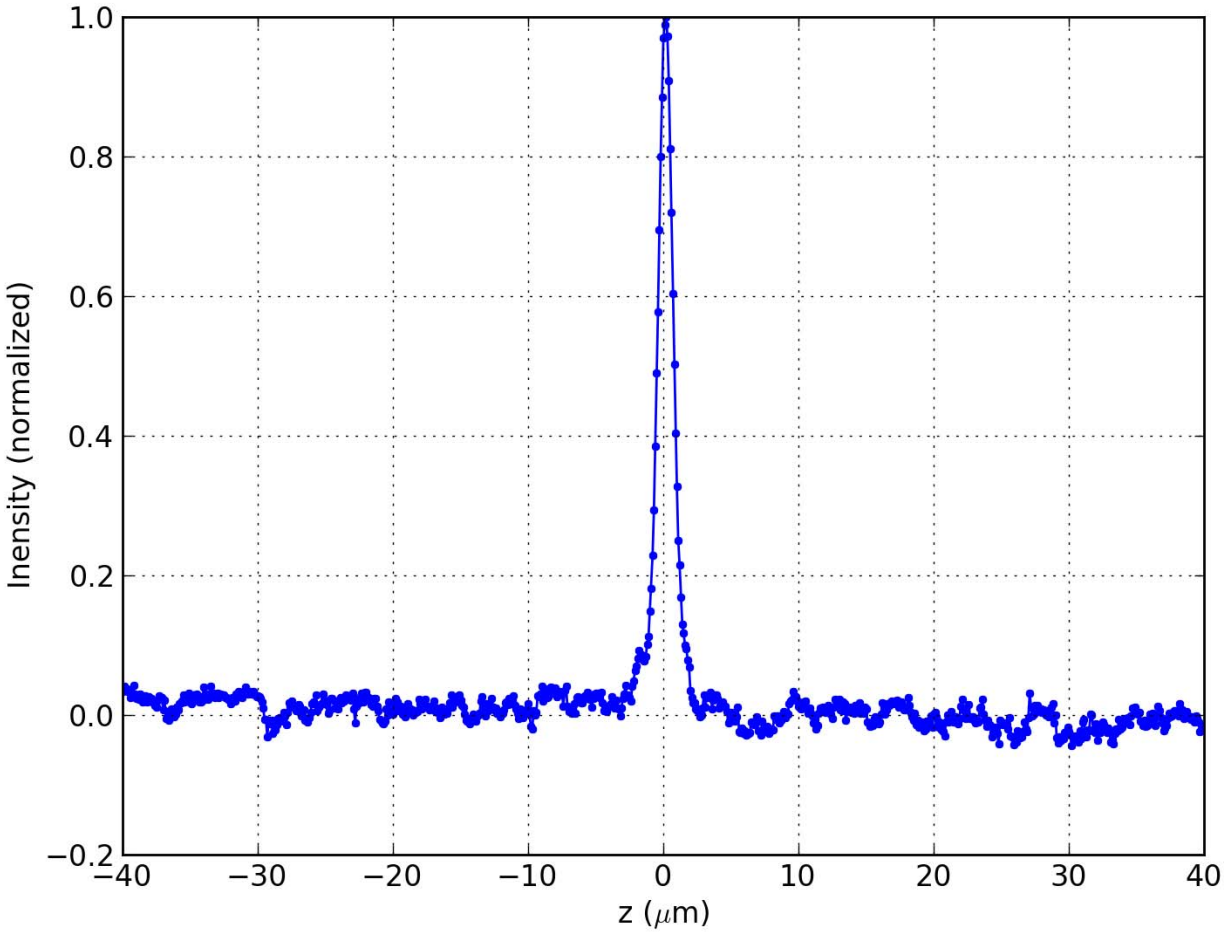
Supplementary Figure 1, Illumination system used for temporal focus/PALM



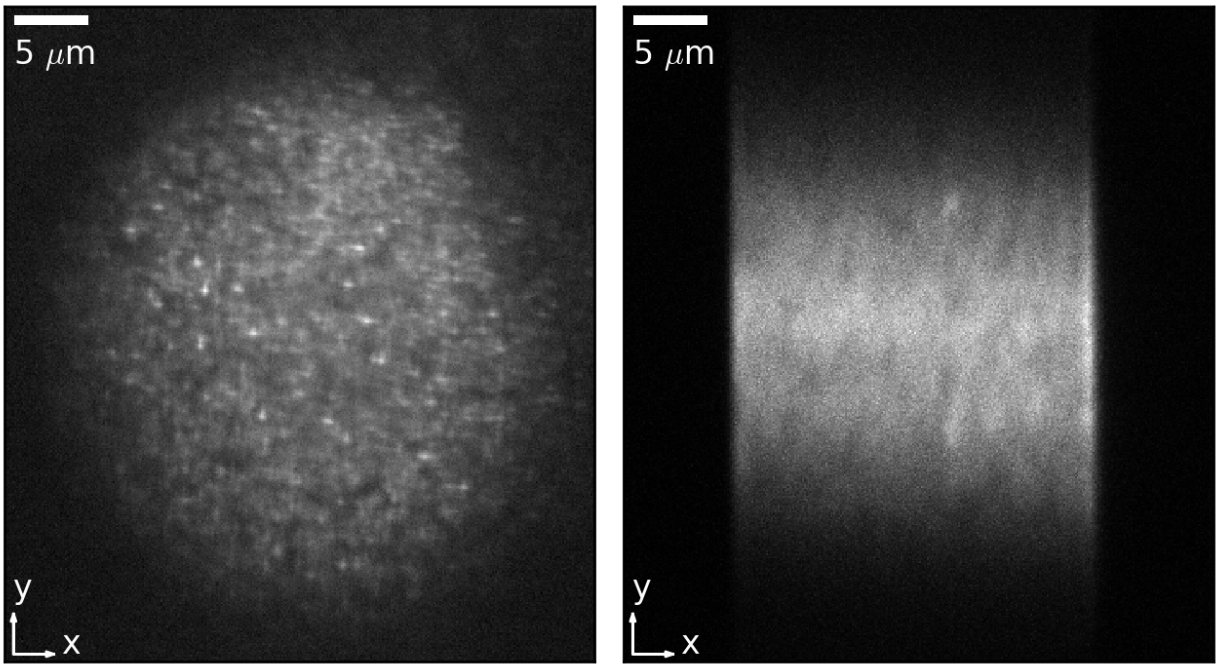
Laser 561 excites the activated form of PA-mCherry1, and laser 405 converts PA-mCherry1 to the activated form. Laser 800 is used for two-photon, temporal focus activation of PA-mCherry1, and may be used either in conventional, nonscanning mode (left beam path) or in a line-scanning mode (highlighted with red beam). The remaining elements, as described in the supporting text, are used either to produce a temporal focus, or to overlap the various lasers beams, adjust their diameters, and ensure their mutual propagation in a common direction. For clarity, illumination is shown only along the 561 nm excitation path (green beam) and the 2 photon line-scanning temporal focus path (red beam). Symbol key: HWP = half wave plate; POL = polarizer; RM = removable mirror; GAL = galvanometric scan mirror; BE = beam expander; CBE = cylindrical beam expander; GR = diffraction grating; DC = dichroic beamsplitter; AOTF = acousto-optic tunable filter; ND = neutral density filter; FL = focusing lens; CL = collimating lens; RL = relay lens. Reflective mirrors are shown as filled in rectangles, and other lenses referred to in the text are shown as ellipsoids with focal lengths as indicated.

Supplementary Figure 2, Sectioning performance of point-scanning two-photon illumination



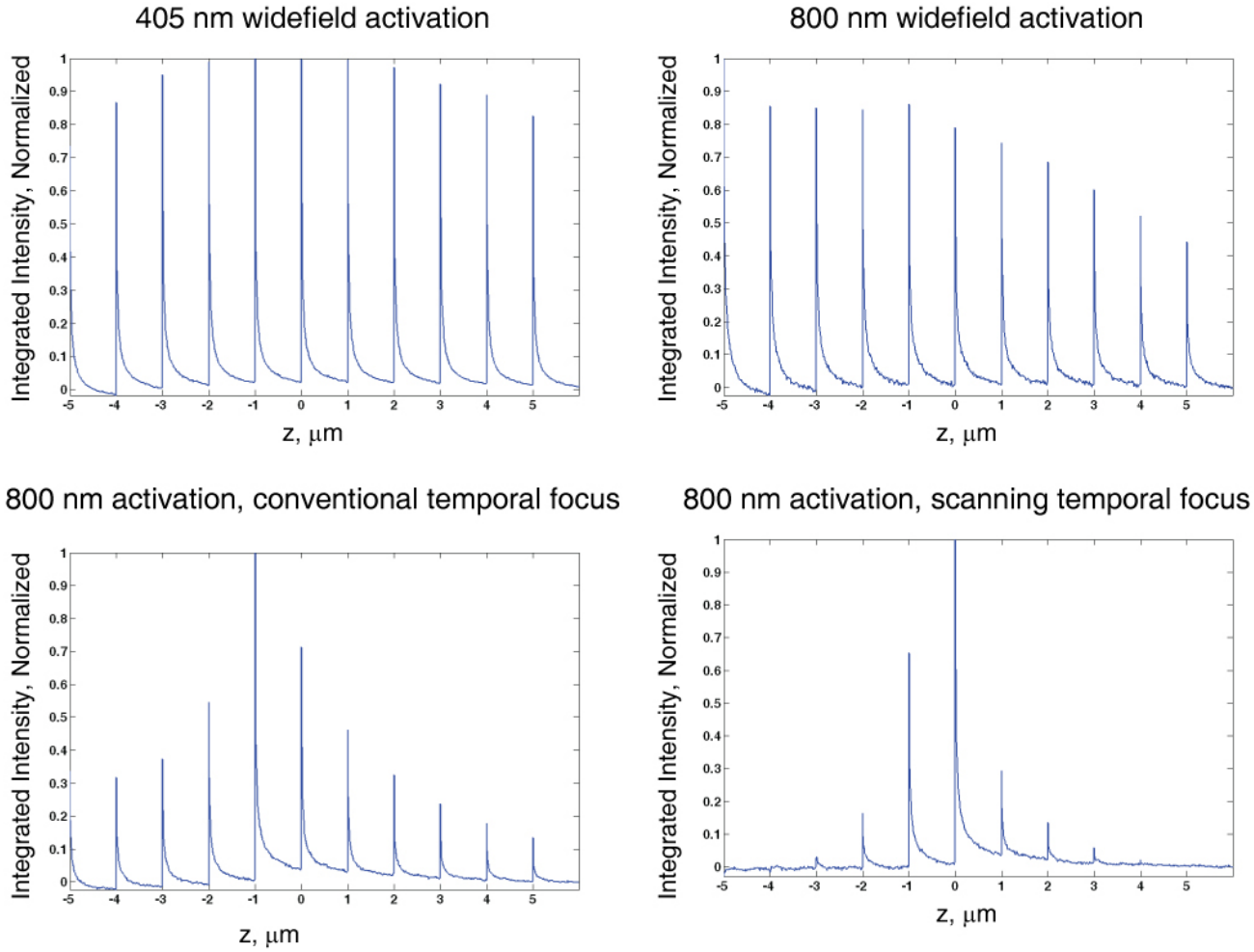
Data were collected on a commercial system (Intelligent Imaging Innovations, Vector) that we tested on our microscope. Integrated axial response of ~800 nm thick quantum dot film, moved through 80 μm axial range in 100 nm steps, and calculated by integrating fluorescence over a 25 μm by 25 μm area. Full width at half maximum (FWHM) is 1.3 μm, comparable to line-scanning temporal focus shown in **Fig. 1c**.

Supplementary Figure 3, Excitation and activation beam profiles



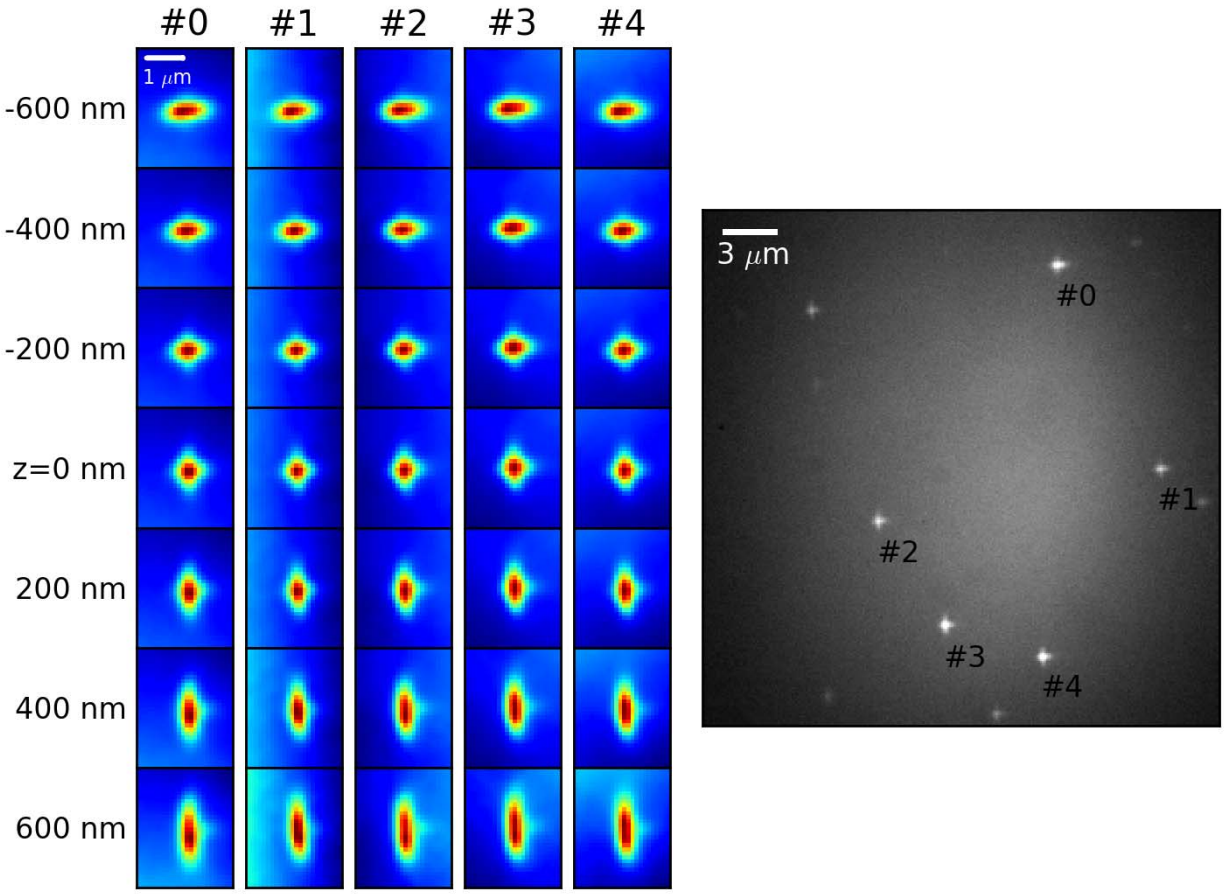
Fluorescence from an 800 nm thick quantum dot film illuminated with 561 nm excitation beam (left), and 800 nm line-scanning temporal focus activation beam (right). The excitation profile shows is roughly Gaussian in both dimensions, whereas the scanned temporal focus intensity is approximately Gaussian in the non-mechanically scanned dimension.

Supplementary Figure 4, Photoactivation of PA-mCherry1 under different imaging modalities



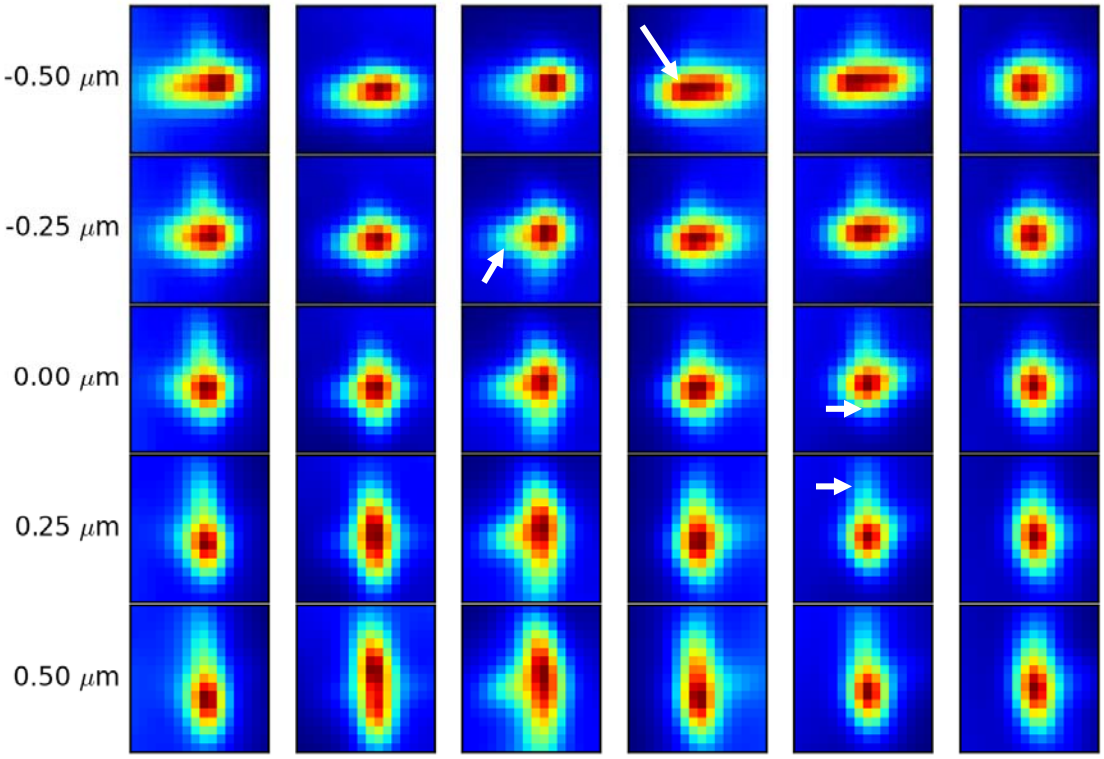
Purified PA-mCherry1 (a gift of George Patterson) was deposited nonspecifically on a glass coverslip, activated as indicated, and excited using 561 nm. Samples were imaged for 100 frames at each z position in order to sufficiently bleach activated molecules before advancing to the next z position. Fluorescence was integrated over a 25 μm by 25 μm area, and normalized by subtracting a baseline fluorescence reading and dividing the result by the maximum number of fluorescence counts. Note that the results here are similar to those reported in **Figure 1c**, with scanning temporal focus producing the most axially confined activation profile. We speculate that the axial broadening compared to **Figure 1c** is due to 561 nm alone induced activation of PA-mCherry1. Asymmetry in the traces is due to bleaching of the sample during the experiment, which we did not correct for.

Supplementary Figure 5, Translational invariance of PSF, taken with cylindrical lens inserted in imaging path



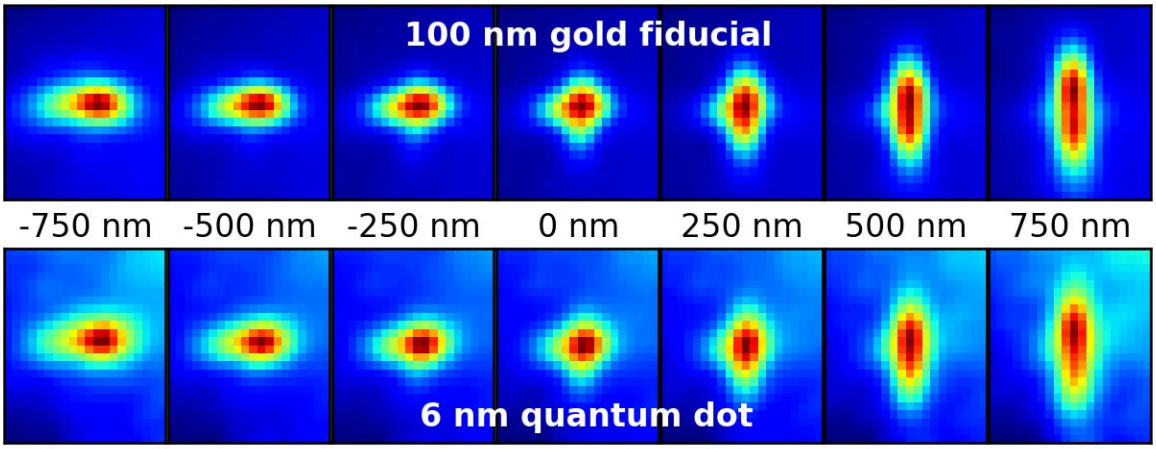
(Left) Axial variation of different 100 nm gold fiducial markers (different columns) at indicated z position, taken on the same coverslip and field of view. (Right) An image of the coverslip at z=0, showing the relative positions of the fiducials. In contrast to **Supplementary Figure 6**, the PSFs of different fiducials are near-identical, supporting the conclusion that PSF variation is sample-dependent, not position-dependent.

Supplementary Figure 6, Sample induced variation of emission PSF, taken with cylindrical lens inserted in imaging path



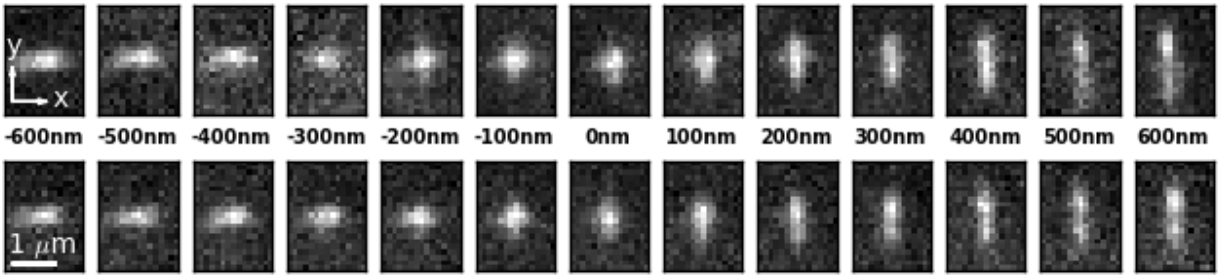
Images are of 100 nm gold fiducial markers at indicated z position, taken on different coverslips (each column represents a calibration stack taken on a different coverslip) with presumably different tilt and thickness. Arrows indicate features that cannot be easily accounted for with a theoretical PSF, such as ‘wings’ in horizontal or vertical directions or asymmetric intensity distribution in the center of the PSF.

Supplementary Figure 7, PSF Comparison between different sized markers.



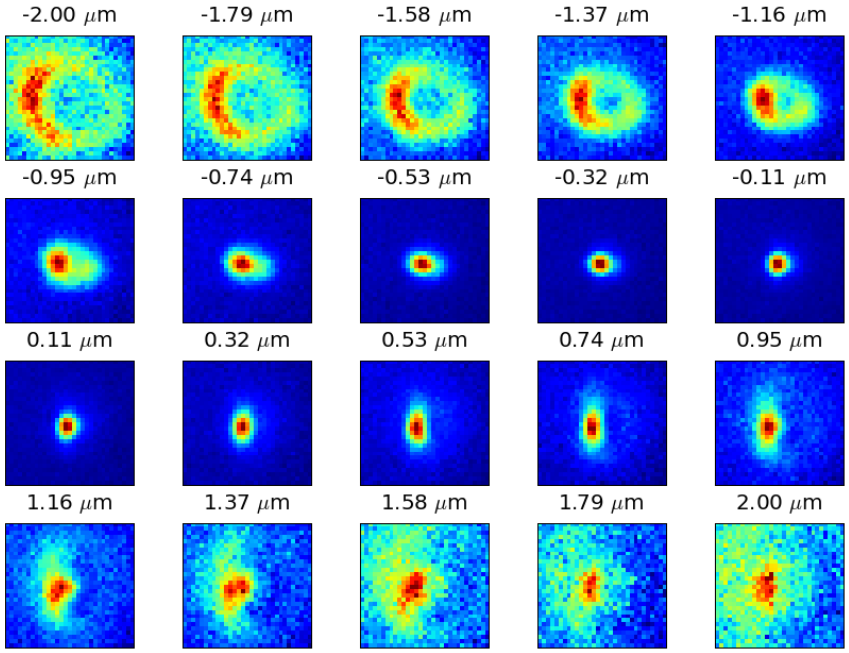
On the same coverslip, and in the same field of view, we compared the PSF of a 100 nm diameter gold fiducial (typically used for 3D localization and drift correction) and the PSF of a ~6nm diameter CdSe/ZnS quantum dot at indicated axial positions. PSFs show remarkable similarity; small differences in background are highlighted by the chosen color map.

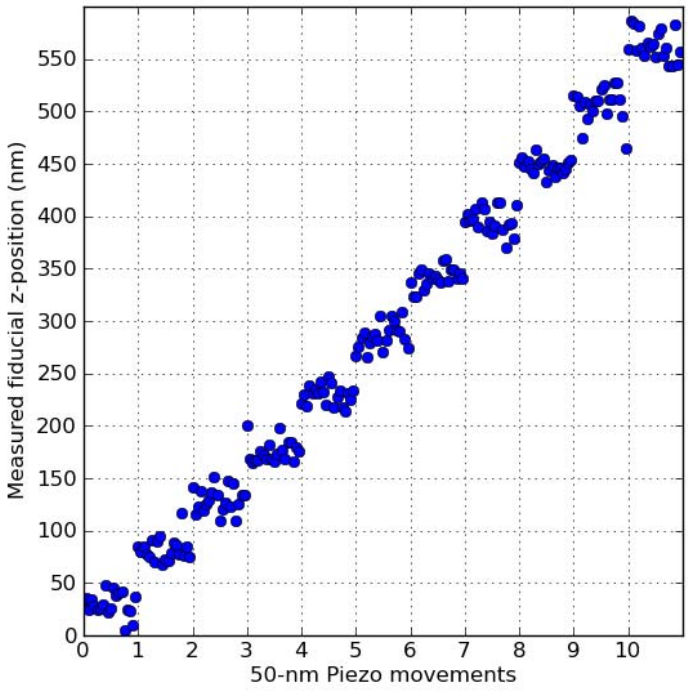
Supplementary Figure 8, PSF comparison between molecules and fiducial



On the same coverslip and in the same field of view, we compare raw images of a 100 nm gold fiducial (bottom row) at known, indicated z-positions to raw images of single PA-mCherry1 molecules at measured z-positions (top row). The brightest molecule localized at each z-position (+/-25 nm) is shown. Because the molecules bleach rapidly, each frame in the top row is of a different molecule. Images are taken from the dataset that produced **Fig. 3**.

Supplementary Figure 9, Model-independent subdiffractive localization code combined with inherent aberrations of imaging optics can resolve 50 nm z steps.

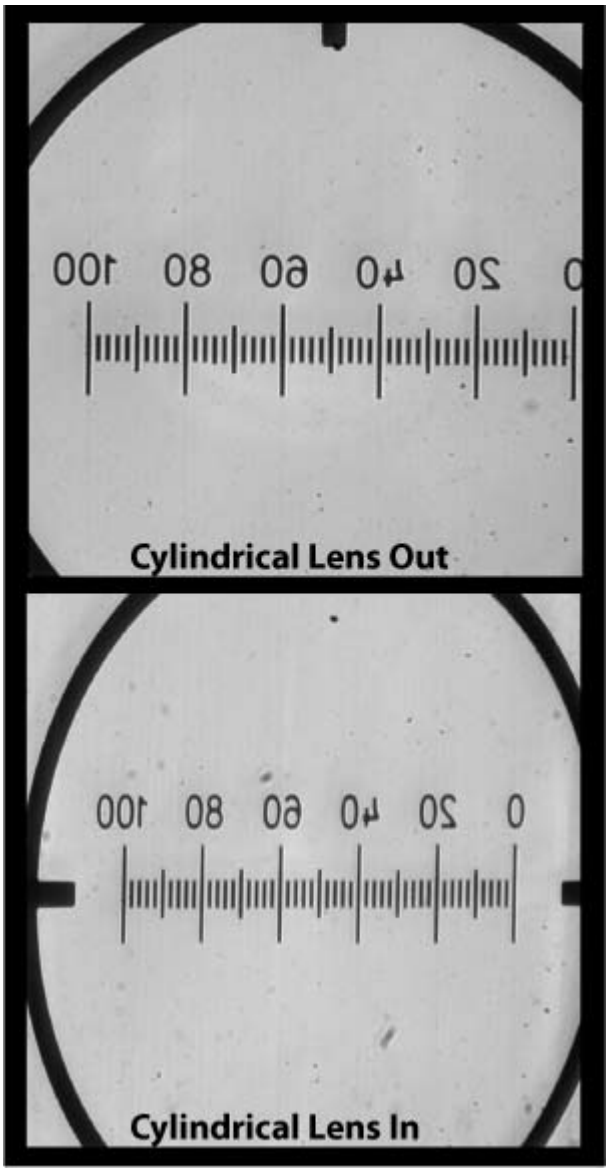




A 100 nm gold fiducial bead was moved in 50 nm axial steps using our z piezo stage and imaged (100 ms integration, EM-gain 300, 20 frames per z position) without a cylindrical lens. Images of

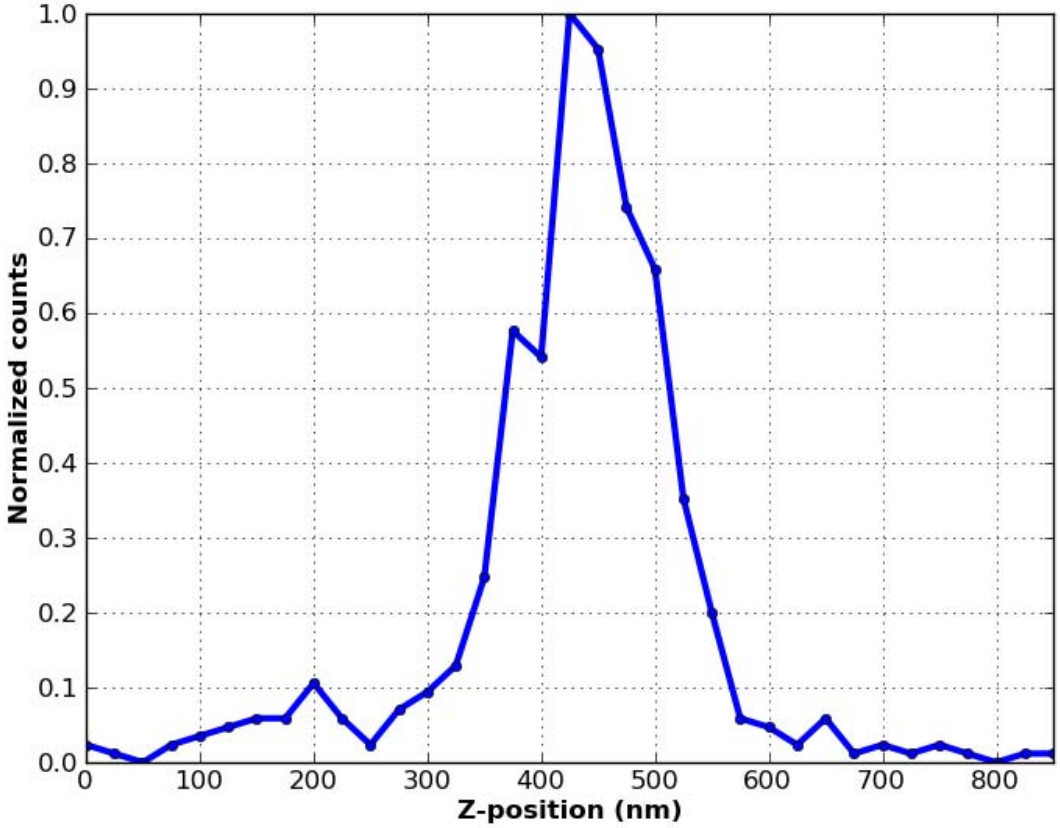
the bead (top) were localized (bottom) using our model-independent code. Steps are clearly resolved, showing that the inherent defocus and aberrations of our imaging optics provide sufficient axial variation for 50 nm z resolution. However, this PSF is not ideal for 3D localization, as at large defocus values, the PSF spreads across many pixels thus degrading the signal-to-noise ratio.

Supplementary Figure 10. Image distortion with a cylindrical lens



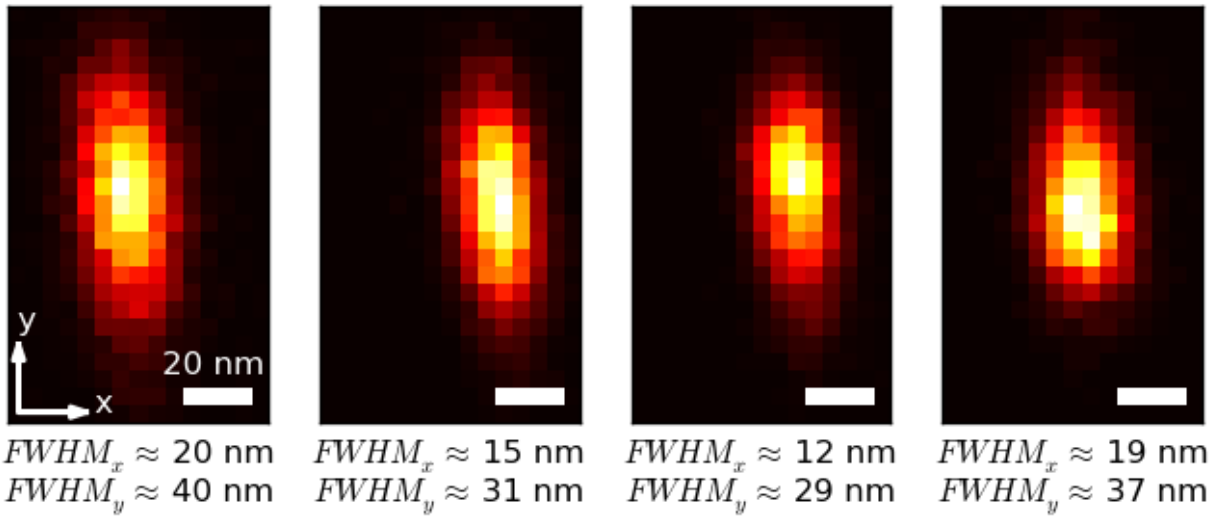
Top: 100 μm reticule viewed through imaging system, without cylindrical lens. Bottom: 100 μm reticule viewed through imaging system, with cylindrical lens placed in the imaging path. Note significant asymmetry between horizontal and vertical dimensions (112 nm/pixel in X, 147 nm/pixel in Y), which can be corrected by a simple linear scaling factor in each direction.

Supplementary Figure 11, Apparent thickness of a PA-mCherry1 protein layer as measured with our 3D PALM localization algorithm



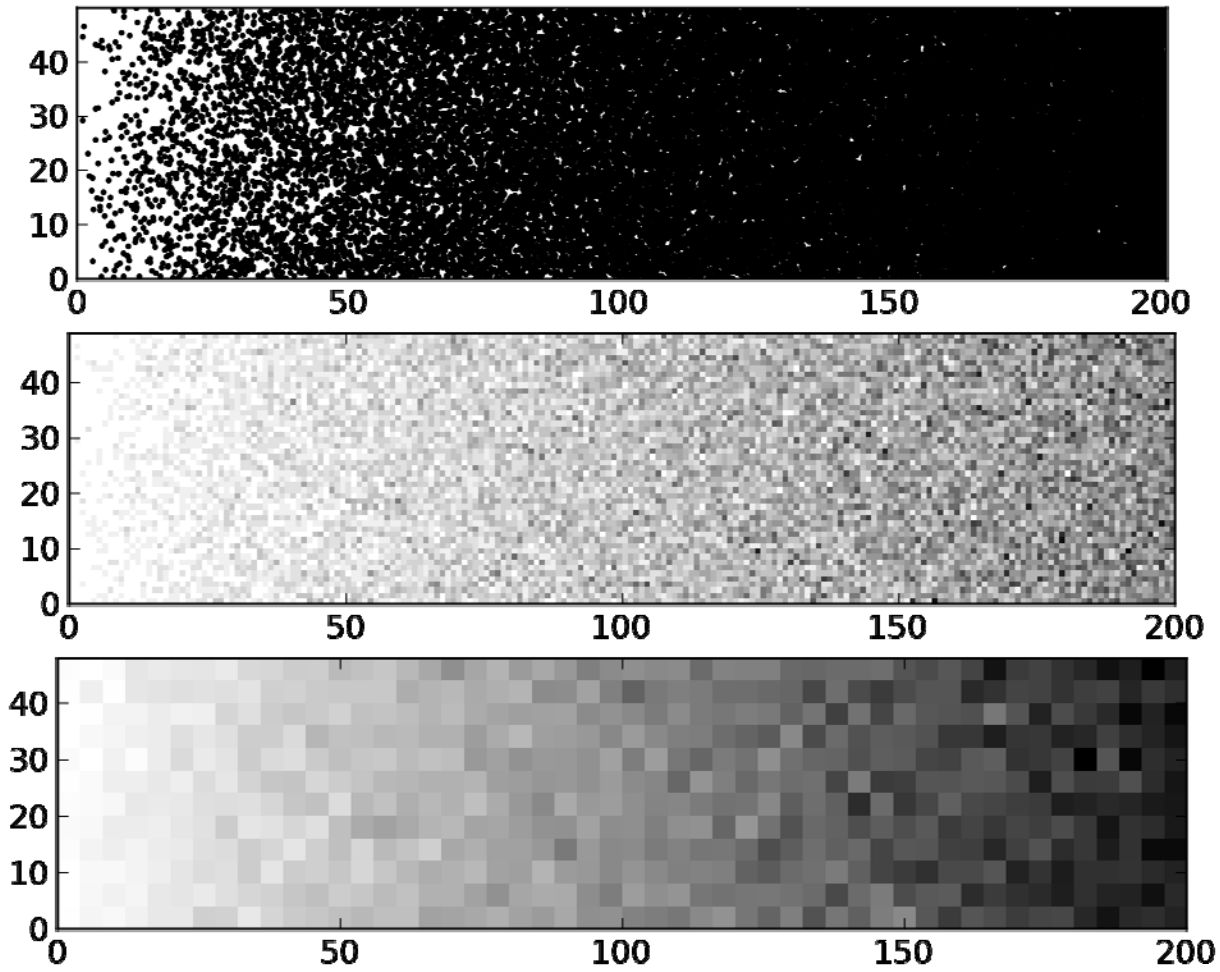
Purified PA-mCherry1 (a gift of George Patterson) was deposited on a clean glass coverslip and 3D PALM- imaged in 1X PBS. The apparent thickness (FWHM) of the layer was calculated from the histogram of unlinked localizations in a 3.7 μm by 4.5 μm area, as ~150 nm. No attempt was made to correct for coverslip tilt or surface roughness, so we expect the calculated thickness to overestimate the true axial resolution of our technique. 536 unlinked localizations with correlation strength > 0.5 were used in generating the histogram.

Supplementary Figure 12, Lateral localization precision for four different gold beads



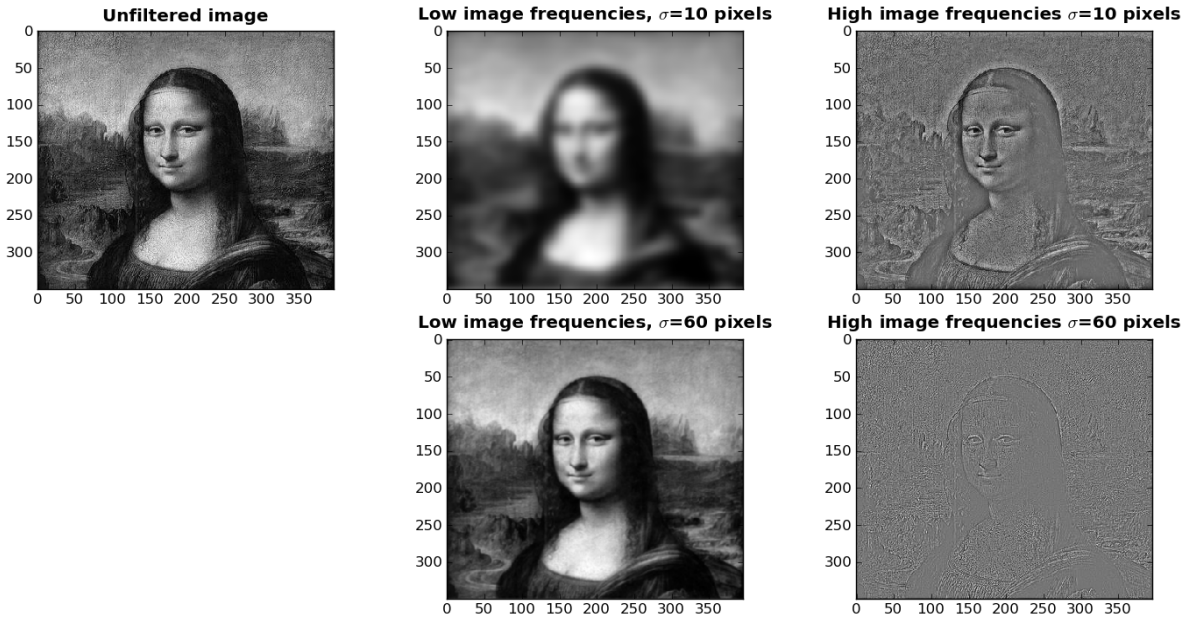
Beads were deposited on a clean glass coverslip and 3D PALM-imaged in 1X PBS for 10000 frames (100 ms exposures). Resulting localizations are shown as 2D histograms with 5x5 nm pixels, and indicated full-width, half-maximum (FWHM) of the resulting distributions in the x- and y-directions are shown below each histogram. These fiducial particles were imaged ~500 nm out-of-focus, such that their images were tall and narrow with an aspect ratio of about 2:1, consistent with the FWHM in the y-direction being about 2x the FWHM in the x-direction. Spread in repeated localizations depends on many factors, including degree of defocus and particle brightness; these fiducials were ~2-3x as bright as a typical PA-mCherry1 molecule. Directly measuring the spread in repeated localizations of a single PA-mCherry1 molecule is difficult due to bleaching.

Supplementary Figure 13, Effect of variable bin size on image histograms



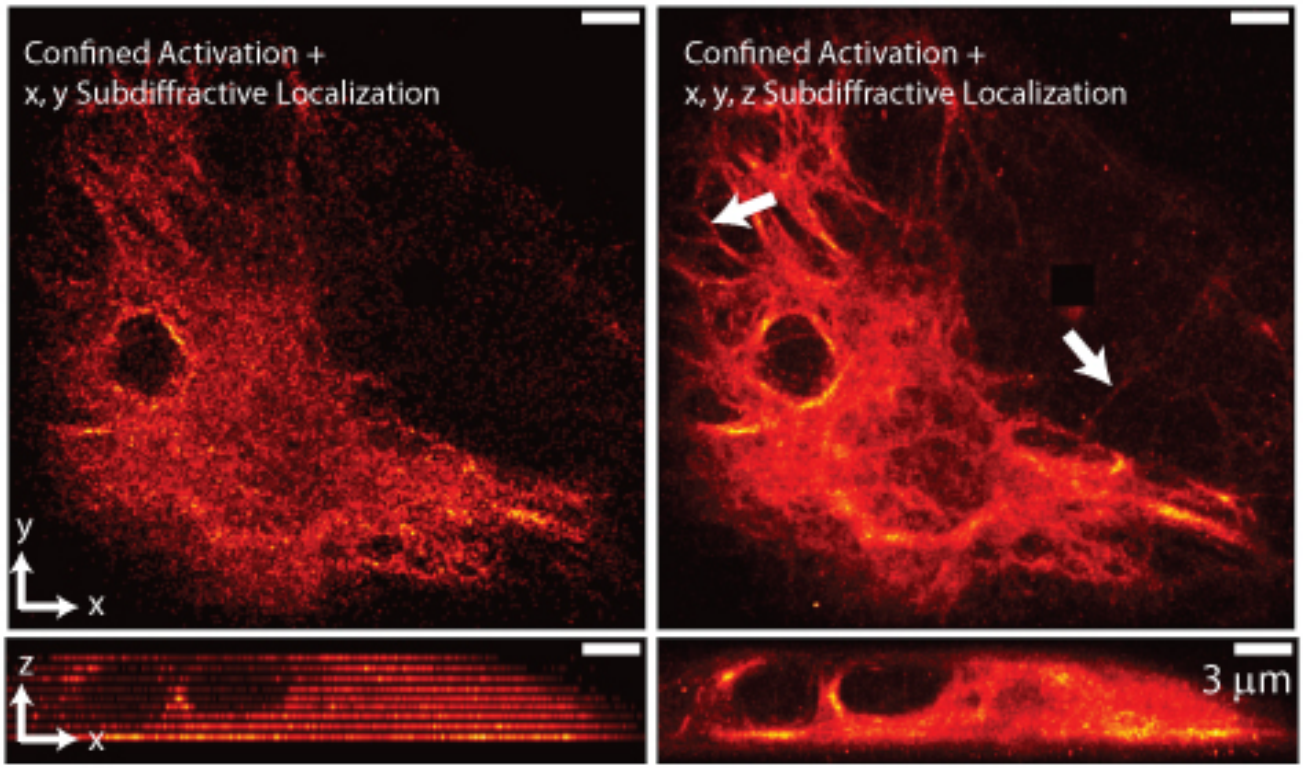
A linear gradient in particle density from left to right is displayed in three ways. Top: A scatterplot accurately conveys particle positions in the low-density region on the left, but becomes useless as density increases to the right. Middle: A histogram with 1x1-unit bins accurately conveys punctate particle positions on the left and gives an impression of increasing density to the right, but the density gradient is difficult to perceive at higher densities. Bottom: A histogram with 4x4-unit bins obscures individual particle positions, but clearly conveys the linear density gradient.

Supplementary Figure 14, Image smoothing



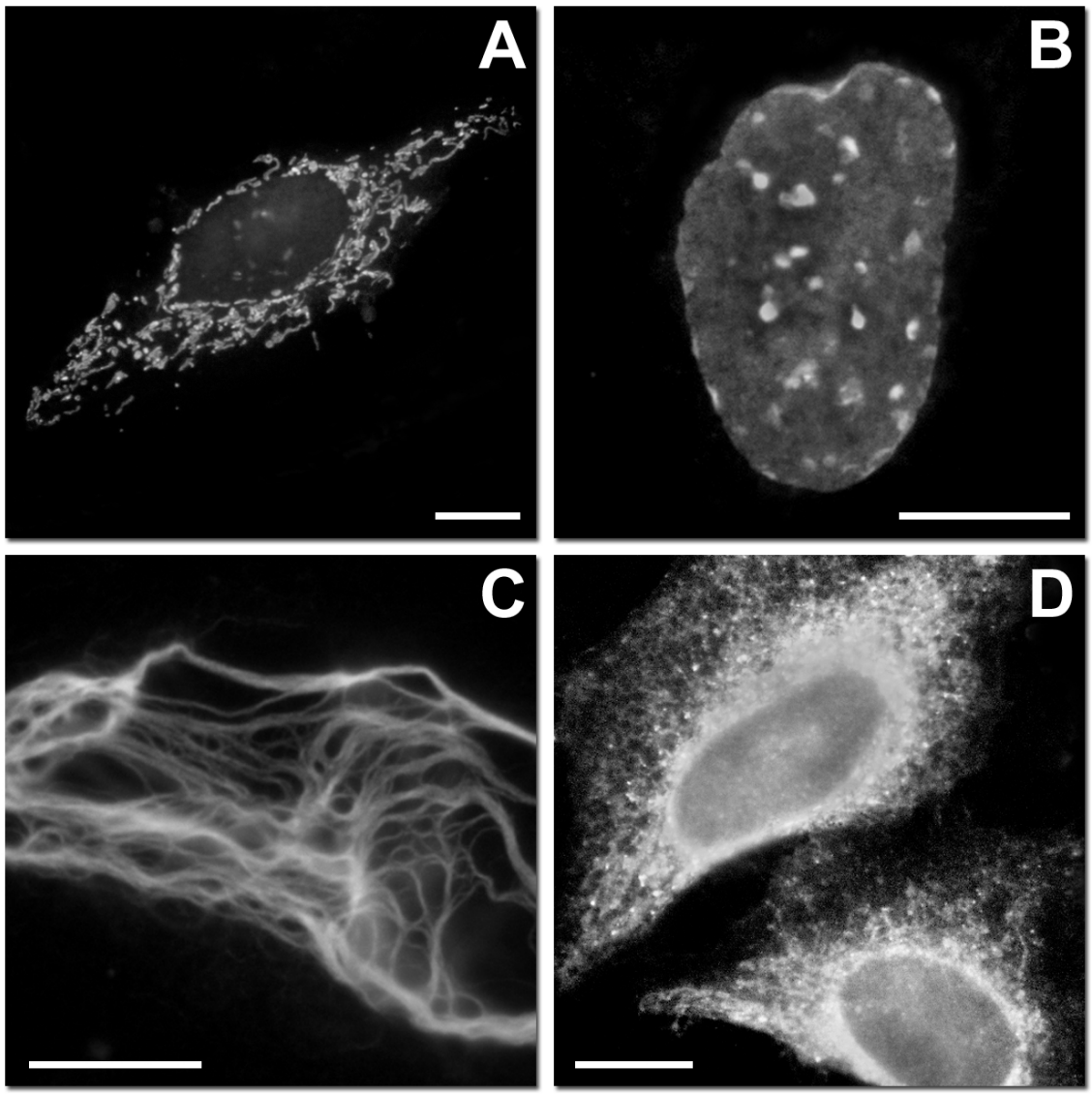
Top row: Unfiltered image, low-frequency image Gaussian-smoothed with σ=10 pixels, and remaining high-frequency image. Removing these high-frequency components clearly degrades the image. Bottom row: Low-frequency image Gaussian-smoothed with σ=60 pixels, and remaining high-frequency image containing mostly noise. Removing these high-frequency components does not significantly degrade the image. Note that σ in this context refers to the width of the cutoff in the frequency domain filter we used; everywhere else in the main text and supplementary information σ refers to a spatial domain blurring.

Supplementary Figure 15, Improvements over 2D subdiffractive localization



Left: A Cos7 cell expressing PA-mCherry1 vimentin, PALM-imaged with confined activation and analyzed with 2D subdiffractive localization methods. Fitting with a 2D Gaussian PSF ignores localizations that result from probe molecules further than a few hundred nanometers from the focal plane, thus producing a 3D data set with clear gaps in the axial direction (bottom panel, xz maximum-intensity projection, 72,000 localizations). Gaps and the low number of localizations also degrade the xy view (top panel). Right: The same dataset, analyzed with 3D subdiffractive localization methods to produce a smoothly connected axial view (bottom panel, xz maximum-intensity projection, 2.3 million localizations). This method does not ignore out-of-focus molecules, providing increased localization density which improves the xy view (top panel, xy maximum-intensity projection). Arrows highlight features in the right projection that are not visible with 2D localization, due to the far fewer number of localizations.

Supplementary Figure 16, Widefield fluorescence imaging of fixed HeLa cells expressing PA-mCherry1 fusion vectors



(A) MTS-7-PA-mCherry1; (B) PA-mCherry1-10-lamin B1; (C) vimentin-7-PA-mCherry1; (D) Calreticulin-5-PA-mCherry1-2-KDEL. Scale bar represents 10 μm.

Supplementary Table 1, Parallels between PALM- and conventional- images

Quantity	**PALM image**	**Conventional image**
Fundamental measurement	Single localization	Single photoelectron
Measurement source	Single fluorescent particle	Single fluorescent particle
Theoretical measurement precision	Unlimited	CCD Pixel Size
Spread in repeated measurements	Photoelectron counting errors	Diffraction
Typical PSF dimension	< 1/10 emission wavelength	< 1 emission wavelength
Typical measurements per PSF	Few	Many

Supplementary Note 1, Parameters chosen to produce diffraction-limited temporal focus

In principle, conventional temporal focus can provide axial sectioning equivalent to standard line-scanning two-photon microscopes, and line-scanning temporal focus can provide sectioning equivalent to standard point-scanning two-photon microscopes. This is unsurprising, as temporal focusing simply replaces mechanical scanning in one direction with passive scanning, trading power density for repetition rate. This trade has no drawback in our setup: standard point-scanning with more than ~10 mW average power damages our samples, but our femtosecond oscillator outputs nearly 4000 mW average power at 800 nm, allowing for reasonable power densities in temporal focusing.

In practice, achieving diffraction-limited sectioning with a temporal focusing microscope requires careful design. We found the most crucial parameter to choose correctly was the magnification between the diffraction grating and the sample. Using a too low a magnification does not give diffraction-limited sectioning, and using too high a magnification overfills the back focal plane (BFP) of the microscope objective, wasting power.

Equation 2 in reference[1] suggests a useful guideline: use magnification such that the instantaneously illuminated area at the diffraction grating is de-magnified to form a diffraction limited spot at the sample,

$$\frac{c\tau}{\sin\alpha} \approx \frac{M\lambda}{2NA}$$

With c the speed of light, τ the pulse width, α the incidence angle onto the grating, M the magnification between grating and sample, λ the wavelength and NA the numerical aperture of the objective.

In our setup, the instantaneously illuminated spot on our diffraction grating (left side of the equation above) is 63 μm. For our 1.2NA objective with 800 nm illumination, a diffraction-limited spot has dimensions ~ 333 nm, suggesting our optimal magnification is ~190X.

We tested sectioning performance for a range of magnifications, and found that sectioning performance improved with increasing magnification up to ~500X. Further increasing our system's magnification above 500X did not noticeably improve sectioning, but decreased the amount of light delivered to the sample. One possible explanation for these observations is that the shape of the instantaneously illuminated spot on the grating may not match the shape of a diffraction-limited spot, and requires further demagnification for optimal sectioning.

Equivalently, producing a diffraction-limited spot requires filling the BFP of the microscope objective uniformly, but a magnification of only ~190X leaves substantial intensity variation across the BFP, which can be smoothed to uniformity by additional magnification.

Because the BFP of the microscope objective is a Fourier plane of the grating's surface, different colors diffracted from the grating focus at different transverse positions in the BFP. The spectrum of our laser pulse is roughly Gaussian in shape, but at a magnification of 500X, only the central portion of our spectrum passes the objective. This portion of the spectrum is fairly uniform in intensity, and we speculate that a laser pulse with similar bandwidth but a more uniform spectral intensity would require less magnification to achieve optimal sectioning.

Supplementary Note 2, Three dimensional model-independent subdiffractive localization

Following data acquisition, we process the data in the Python programming language[2,3]. Data processing steps are summarized as follows:

1. Construct calibration stack
2. Identify candidate particles in each data image
3. Localize each candidate particle using calibration stack
4. Correct for drift
5. Optionally, link localizations, and re-localize
6. Construct image histograms from localization data

In more detail,

1. Construct calibration stack

Crop the calibration image stack in the xy dimensions to contain a single fiducial and to minimize blank space around the fiducial. Average all cropped calibration images taken at the same piezo position. Compare the first z series to subsequent repetitions to ensure that sample drift is not substantial during the calibration acquisition. Smooth the calibration stack in 3d with a Gaussian filter, σ = 1 pixel in xy and 5 pixels in z.

2. Identify candidate particles in each data image:

Apply a Gaussian-Laplace filter with σ = 4 pixels to each data image, removing slowly-varying background and quickly-varying noise. Select candidate particles from the filtered image as follows:

a.) Mark all pixels in the filtered image that are above a user-specified threshold (typically 4 standard deviations of the filtered image above the filtered image mean).

b.) Place a rectangle (with the same xy dimensions as the calibration stack) centered on the brightest marked spot in the filtered image. Record the rectangle's location, and unmark all the pixels inside this rectangle.

c.) Repeat b.), placing rectangles, unmarking pixels, and recording rectangle locations until no pixels in the image are marked. All coordinates recorded are candidate 'bright' particles.

d.) Construct the differential filtered image by subtracting the previous filtered image from the current filtered image, and repeat the candidate selection process (steps a.)-c.)) to find candidate 'birth' particles, regions where brightness increased substantially from the previous frame.
e.) Finally, repeat candidate selection (steps a.)-c.)) on the negative of the filtered differential image to find candidate 'death' particles, regions that decreased in brightness substantially from the previous frame.

For each dataset, the 'bright' detection threshold and the 'birth'/'death' detection threshold were independently adjusted (typically from 3-5 standard deviations of the filtered image) to minimize obvious false positives and false negatives. False positives cause unnecessary computation, and false negatives cost potential localizations.

3. Localize each candidate particle using calibration stack:

For each candidate, load the xy-subregion of the raw data image recorded during candidate selection. For 'birth' candidates, subtract the previous frame from the current frame. For 'death' candidates, subtract the current frame from the previous frame. Cross-correlation is then used to find the subpixel x, y, z shift which will best align the candidate image with the calibration stack. This operation is conceptually simple:
a.) Bandpass filter the calibration stack and the data image, to remove slowly-varying background and quickly-varying noise. Normalize the filtered images so that cross-correlations will vary between -1 (perfect anticorrelation) and 1 (perfect matching correlation). In practice, a correlation of less than 0.2 was rare, and correlation strengths above 0.4 typically indicated an acceptable match between calibration images and data.
b.) Shift the calibration stack (with periodic boundaries) in x, y, and z by some amount (x_s, y_y, z_s). Use interpolation if necessary to calculate sub-pixel shifts.
c.) Pointwise-multiply the data image by the first xy-slice of the shifted calibration stack, and sum the result to produce a correlation strength.
d.) Find the shift (x_m, y_m, z_m) which maximizes correlation strength.
e.) Add this shift to the xy position of the candidate selection rectangle to give a localization coordinate.
f.) Record the localization coordinate and continue with steps c.)-e.) for the next candidate molecule.

This process must be done with great care to ensure useful results in reasonable computation time. Others[4] have described how to greatly accelerate a similar (but two-dimensional) search using Fourier cross-correlation, Fourier interpolation, and matrix-multiplication-based inverse Fourier transforms. Similar techniques are used here for speed; note that the detailed steps below are equivalent to the simple process (steps a.)-f.)) described above:

a.) Construct a 2d Fourier bandpass filter W for detrending and smoothing (typically the pointwise product of a Gaussian high-pass filter with $\sigma = 1$ pixel and a Gaussian low-pass filter with $\sigma = 7$ pixels). b.) Discrete Fourier transform the calibration stack in the x and y directions using the Fast Fourier transform (FFT). Pointwise-multiply each xy slice by the bandpass filter W from a.). Normalize the resulting slices to have unit vector magnitude (the sum of the squares of the filtered image pixel values is one). Fourier transform this array in the z-direction, and store it, since it is used many times.

c.) FFT the candidate image, multiply by the bandpass filter from a.), and normalize as in b.). Tile copies of this array in the z-direction to give an array with the same dimensions as the calibration stack.

d.) Pointwise-Multiply the FFT'd, filtered, normalized calibration stack by the FFT'd, filtered, normalized, z-repeated data image. Mathematically, for each image I, calculate C:

$$C = \text{Tile}(\text{Normalize}(W * \text{FFT}_{XY}(I))) * \text{FFT}_{Z}(\text{Normalize}(\text{Tile}(W) * \text{FFT}_{XY}(\text{Calibration})))$$

where W is the bandpass filter described in a.), '*' denotes pointwise multiplication, the 'Normalize' operation is described in b.), and the 'Tile' operation is described in c.). The inverse FFT of array C gives the cross-correlation of the data image with the calibration stack for integer-pixel x, y, and z shifts. If the array C were zero-padded appropriately, its inverse FFT would give the same cross-correlation for subpixel shifts, but this operation becomes very expensive for small shifts.

e.) Instead, use a series of inverse Fourier transforms similar to the process described in reference[4] to find the maximum of the cross correlation. The first transform truncates C by a factor of 2 to 4 before performing an inverse FFT, giving a coarse, inexpensive cross-correlation of 2 to 4 pixel shifts. The second transform zero-pads C by a factor of 2, but uses a matrix-multiplication inverse discrete Fourier transform to calculate half-pixel-shift cross-correlations in

an 11x11x41 xyz-shift region centered on the maximum of the first, coarse cross-correlations. The third and final transform zero-pads C by a factor of 20 before a matrix-multiplication-based inverse Fourier transform to calculate 1/20th-pixel-shift cross-correlations in an 11x11x41 xyz-shift region centered on the maximum of the second, half-pixel cross-correlation.

g.) Record the subpixel location of the maximum of our cross-correlation. Add this shift to the xy position of the candidate selection rectangle to give a localization coordinate.

f.) Record the localization coordinate and repeat steps c.)-e.) of this process with the next candidate molecule.

4. Correct for drift:

After localization, examine a plot of x, y, z localizations and correlation strength vs. image number, and specify a filter based on localization and correlation strength that uniquely identifies a fiducial. Use piecewise linear interpolation followed by user-specified gaussian smoothing to estimate the sample drift versus time.

5. Optionally, link localizations, and re-localize:

A single fluorescent molecule may produce several candidate particles spread out over several frames, and often reasonable guesses can be made to link multiple candidates. Starting with each 'birth' localization, search the next three frames for a subsequent 'death' localization within 3 pixels in x and y, and 10 pixels in z. If no 'death' is found but a 'bright' localization is found which matches the 'birth', search an additional three frames for a 'death' localization. For each 'birth', 'death' pair linked this way, construct an image which is the average of all frames between the 'birth' and 'death'. Subtract the average of the frame before the birth and the frame after the death (background), and relocalize the result to generate a linked, background subtracted localization. These localizations may be further analyzed for quantities such as total number of photoelectrons. Linked localizations generally have a higher signal-to-noise and thus higher localization precision than unlinked localizations, but linking also leads to a decreased number of total localizations in the final PALM image. Since resolution is often more limited by number of localizations than by localization precision, linking is not always beneficial.

6. Construct image histograms from localization data

Construct a 3D histogram of xyz localization coordinates. Each xyz bin records the number of localizations in a given xyz range. The localization coordinates are corrected for sample drift and known z-piezo position. Proper choice of xyz bin size is critical. A very small bin size produces a 3D scatterplot of xyz localizations. A scatterplot is a very accurate representation of the measured localizations, but a crowded scatterplot can be difficult to interpret by eye. A larger bin size produces a density map, but too large a bin size degrades image resolution. More discussion is provided in **Supplementary Note 3**.

The user can also specify filters to reject localizations. Typically localizations with a low correlation strength (0.4 or less), and/or that were very out-of-focus (more than 1.5 microns) and therefore had large PSFs with poor signal to noise were rejected.

Supplementary Note 3, Image rendering

PALM records a set of particle positions, often with some estimate of each measurement's error. A 2D or 3D scatterplot with error bars communicates this information completely, but scatterplots become crowded and confusing with increasing measurement density. A variety of visualization schemes have been used for 2D and 3D PALM data, but no clear visualization standard is established. As a result, it is often unclear exactly what a PALM 'image' displays.

There are strong parallels between conventional CCD-based fluorescence microscopy and PALM, and understanding these parallels helps us benefit from established practices in conventional image rendering. In particular, we find the concept of a "PALM PSF" useful: Just as repeated measurements of photoelectron positions from a constant intensity, isolated fluorescent point emitter give the point-spread-function (PSF) of a conventional microscope, repeated measurements of localization positions from a constant intensity, isolated fluorescent point source give the PSF of a PALM microscope. Our definition of a PALM PSF is equivalent to the conventional definition of localization precision; any model that accurately predicts the localization precision must predict the same spread as that observed in repeated localizations. We note that there are many parallels between conventional fluorescence imaging and PALM imaging, summarized in **Supplementary Table 1**.

PALM data is very similar to a conventional fluorescent image, with a smaller PSF volume but typically many fewer measurements per PSF volume. Data recorded from a conventional fluorescence microscope represents photoelectron counts as a histogram, displaying captured fluorescence intensity vs. position. Similarly, a PALM dataset which is too dense to show clearly as a scatterplot with error bars can be represented as a histogram that displays localization density vs. position. Plotting the PALM image this way abandons the error bars, just as a conventional fluorescence image does. In both types of image histograms, bin-to-bin fluctuations substantially smaller than the measurement PSF cannot be due to structures in the sample, and these artifacts can be safely removed by smoothing.

We find a bin size of ~60x60x60 nm for large 3D histograms (~30x30x5 microns), and ~25x25x25 nm for histograms of smaller subregions is appropriate given our measured localization density. Much larger bins degrade imaging resolution. The effect of a bin size that is

too small is more subtle: while preserving resolution, smaller bins lead to a punctate image that is visually confusing at high densities (**Supplementary Figure 13**).

Slight Gaussian smoothing (σ = 0.4 pixels) is applied to further diminish the effects of counting noise, and we only use localizations within +/- 1.5 μm of imaging focus, localized with a correlation strength of 0.4 or higher to select for more precise localizations. We note that the choice of appropriate image smoothing is qualitative to some degree. We found a useful rule of thumb for image smoothing was to examine both the smoothed image and the removed high-spatial-frequencies, as shown in **Supplementary Figure 14**. When the removed high-spatial-frequencies do not show a recognizable image, their removal is unlikely to degrade image resolution significantly.

References

1 Oron, D., Tal, E. & Silberberg, Y. Scanningless depth-resolved microscopy. *Opt. Express* **13**, 1468-1476 (2005).

2 Oliphant, T. Python for Scientific Computing. *Comput. Sci. Eng.* **9**, 10-20 (2007).

3 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90-95 (2007).

4 Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156-158 (2008).