

Supplementary Materials

Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells

Yuan-I Chen¹, Yin-Jui Chang¹, Shih-Chu Liao², Trung Duc Nguyen¹, Jianchen Yang¹, Yu-An Kuo¹, Soonwoo Hong¹, Yen-Liang Liu^{3,4}, H. Grady Rylander III¹, Samantha R. Santacruz^{1,5,6}, Thomas E. Yankeelov^{1,7,8,9,10,11} and Hsin-Chih Yeh^{1,12*}

¹Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

²ISS, Inc., 1602 Newton Drive, Champaign, IL 61822, USA

³Master Program for Biomedical Engineering, China Medical University, Taichung 406040, TAIWAN

⁴Research Center for Cancer Biology, China Medical University, Taichung 406040, TAIWAN

⁵Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA

⁶Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

⁷Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712, USA

⁸Department of Diagnostic Medicine, The University of Texas at Austin, Austin, TX 78712, USA

⁹Department of Oncology, The University of Texas at Austin, Austin, TX 78712, USA

¹⁰Livestrong Cancer Institutes, The University of Texas at Austin, Austin, TX 78712, USA

¹¹Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

¹²Texas Materials Institute, The University of Texas at Austin, Austin, TX 78712, USA

Supplementary Methods

Supplementary Figures

Supplementary Fig. 1 Generation of the simulated decay histogram with lifetime parameters and photon counts by MC simulation

Supplementary Fig. 2 Schematics of simulated training dataset and training procedures for *flimGANE*

Supplementary Fig. 3 Generator (*G*), a subnet from *flimGANE*, can transform low-photon-count decay histogram into high-photon-count one

Supplementary Fig. 4 *flimGANE* outperformed other analysis methods in FLIM images reconstruction at various photon-count conditions in-silico

Supplementary Fig. 5 The performance of different analysis methods was evaluated by MSE and SSIM

Supplementary Fig. 6 The performance of different analysis methods was evaluated by quality index (QI)

Supplementary Fig. 7 Difference of center of mass (CoM) between the IRF and decay histogram was demonstrated positively correlated

Supplementary Fig. 8 Calibration line for center of mass evaluation (CoME) was consistent over different photon counts

Supplementary Fig. 9 A real-world application of the CoME

Supplementary Fig. 10 *flimGANE* can accurately discriminate fluorescence lifetime barcode beads individually at low-photon-count condition

Supplementary Fig. 11 Two dimensional (lifetime versus intensity) scattered plots from different methods illustrated the success of *flimGANE* for identifying three populations of barcodes with nearly the same size

Supplementary Fig. 12 Performance comparison between *TD_MLE*, *MLE-pattern matching*, and *flimGANE* on fluorescence barcode classification

Supplementary Fig. 13 CFP-g-YFP-transfected cells were selected from fluorescence image by ImageJ (ROI Manager)

Supplementary Fig. 14 Normalized CFP lifetime distributions of CFP-g-TFP MDA-MB-231 cells and Gaussian fitting of each lifetime value for *flimGANE*

Supplementary Fig. 15 The quantification of CFP-g-YFP-transfected MDA-MB-231 cell FLIM images for YFP channel

Supplementary Fig. 16 Comparison of WGAN and GAN on learning to reconstruct the ground-truth mimicking fluorescence decay histograms

Supplementary Fig. 17 Comparison of *TD_MLE* tested on simulated decay histogram or on G_{output} (*TD_MLE_{G_output}*) and *flimGANE*

Supplementary Fig. 18 Incorrect Convallaria FLIM image was reconstructed by *flimGANE* with incorrect IRF input

Supplementary Fig. 19 Influence of fluctuating IRF on *flimGANE* performance

Supplementary Fig. 20 Quality of estimate (*G*-quality score) in *flimGANE* based on known samples

Supplementary Fig. 21 Comparison on discriminability for *TD_LSE*, *TD_MLE*, *DFD_LSE*, and *flimGANE*

Supplementary Tables

Supplementary Table 1 Training database for different applications

Supplementary Table 2 Summary of different photon-count conditions

Supplementary Table 3 Mean squared error (MSE) comparison of different analysis methods in silico

Supplementary Table 4 Peak signal-to-noise ratio (PSNR) comparison of different analysis methods in silico

Supplementary Table 5 Structural similarity index (SSIM) comparison of different analysis methods in silico

Supplementary Table 6 Goodness-of-fit (χ^2) comparison of different analysis methods in silico

Supplementary Table 7 Execution time comparison of different analysis methods in silico

Supplementary Table 8 Summary of different analysis methods for the mixture of two fluorescent dyes

Supplementary Table 9 Summary of each detected bead

Supplementary Table 10 MSE, PSNR, SSIM, and VIF comparison of different analysis methods for Convallaria FLIM images (Standard: medium-count *TD_MLE* FLIM)

Supplementary Table 11 MSE, PSNR, SSIM, and VIF comparison of different analysis methods for live HeLa cell FLIM images (Standard: medium-high-count FLIM)

Supplementary Table 12 Apparent lifetime of CFP calculated with Gaussian distribution fitting

Supplementary Table 13 MSE, PSNR, and SSIM comparison of different analysis methods for FRET FLIM images
(Reference: TD_MLE FLIM)

Supplementary Table 14 MSE, PSNR, and SSIM comparison of different analysis methods for autofluorescence
FLIM images (Reference: TD_MLE FLIM)

Supplementary Methods

Generative adversarial network structure and training. In this work, we trained a deep neural network using a generative adversarial network (GAN)¹. GAN consisted of two subnets, generator (G) and a discriminator (D), which were trained via an adversarial process. G mimicked the ground-truth fluorescence decay histogram while D returns an adversarial loss to the G -output fluorescence decay. The cost functions for the gradient descent update of generator (G_{cost}) and discriminator (D_{cost}) were designed as follows, respectively.

$$G_{cost} = \frac{1}{n} \sum_{i=1}^n \log(1 - D(G(z_i))) \quad (1)$$

$$D_{cost} = \frac{1}{n} \sum_{i=1}^n (\log(D(x_i)) + \log(1 - D(G(z_i)))) \quad (2)$$

where z_i represents the normalized low-photon-count fluorescence decay histogram, and x_i is the normalized ground-truth fluorescence decay histogram. $G(z)$ is the normalized ground-truth mimicking histogram (G_{output}), and $D(x)$ represents the probability that x came from the ground-truth fluorescence decay histogram rather than G_{output} .

While GAN had shown great success, there exist several problems, including the difficulty to achieve Nash equilibrium² vanish gradient due to the use of loss functions in equation³ (1, 2). Wasserstein generative adversarial network (WGAN)⁴ was used to improve the G performance. Wasserstein distance instead of Jensen-Shannon divergence was used to train GAN by evaluating the distance of the distributions between the real data and the generated data. The improved training process could offer strong enough gradients to train the generator than that of the original GAN. The cost functions of WGAN for the gradient descent update of G and D were as follows,

$$G_{cost} = -\frac{1}{n} \sum_{i=1}^n f(G(z_i)) \quad (3)$$

$$D_{cost} = \frac{1}{n} \sum_{i=1}^n (f(x_i) - f(G(z_i))) \quad (4)$$

$$|f(x_1) - f(x_2)| \leq |x_1 - x_2| \quad (5)$$

Where $f(x)$ is a 1-Lipshitz function which is defined in equation (5). To approximate the function f for Wasserstein distance calculation, we trained D parameterized with weights lying in a compact space determined by clipping parameters. Higher value of D output refers to the ground-truth data; while lower value refers to the low-photon-count histogram. This modification of loss function helps to stabilize the training schedule and ensure that training process can lead the deep learning model to converging.

Generative model (G). Residual neural network⁵ and convolution neural network are artificial neural networks (ANNs) architecture that were first proposed for image recognition, easing the training of networks that are substantially deeper than those used previously. A similar network architecture has also been successfully applied in recent image reconstruction and image synthesis applications⁶. The structure of the generative network used in this work is demonstrated in **Fig. 1**, which consists of a convolutional block, a multi-task layer ($n = 3$), a decoding

layer explicitly, and a residual block implicitly. Inside the residual block, the convolutional block consists of two convolutional layers and pooling layers followed by a flatten layer, within which it performs

$$y = \text{ReLU}[\text{Conv}\{\text{ReLU}[\text{Conv}\{\text{Concat}(x_{1:256}, x_{256:512})\}]\}] \quad (6)$$

where x represents the input of the generative model, normalized low-photon-count fluorescence decay histogram and corresponding instrument response function (IRF), y is the output of the convolutional block. $\text{Concat}()$ is the concatenation operation of two inputs. $\text{Conv}\{\}$ is the convolution operation, $\text{ReLU}[]$ is the rectified linear unit activation function,

$$\text{ReLU}[x] = x^+ = \max(0, x) \quad (7)$$

The dimension of the output of each ReLU activation function is reduced by AveragePooling layer. Then a multi-task neural network with hard parameter sharing converts the high dimensional flattened output into three tasks, and each task corresponds to each lifetime parameter (for example, bi-exponential decay model). The last network, decoding layer, termed multilayer perceptron with the activation functions as $\tanh()$ to force the range of the output to lie within -1 and 1, maps the 3 tasks into 256 channels of output that corresponds to the fluorescence decay histogram. Instead of learning a direct mapping toward ground-truth fluorescence decay histogram, we reframe the process with the residual learning framework by introducing a residual connection between one of the inputs, normalized low-photon-count decay histogram, and the model's output.

Discriminative model (D). The structure of the discriminative model begins with a densely connected neural network with 128 nodes. Then the output is fed into other densely-connected neural networks with 64, 8, and 1 node. All layers instead of the last one has a sigmoid activation function, which forces the output to lie within the range from 0 to 1, defined as follows,

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (8)$$

The last layer has the linear activation function to output the score corresponding to the input histogram fed into D .

Estimator model (E). On top of the components in the GAN framework, here we also constructed an estimator that focused on predicting the lifetime parameters of interest from the high-photon-count fluorescence decay histogram. In *flimGANE* analysis, the output of the generator, normalized ground-truth mimicking fluorescence decay histogram, would be fed into the estimator model, E , to infer the lifetime values. This holds a great potential to improving the predictive power of the lifetime for more than two species. The structure of the estimator model begins with two densely connected neural networks with 64, and 128 nodes for incoming IRF and high-photon-count fluorescence decay histogram, respectively. These two outputs are concatenated by a concatenation layer. The output of the concatenation layer is fed into the multi-task neural network ($n = 3$) with hard parameter sharing, and a multilayer perceptron with a single hidden layer, whose output is the corresponding fluorescence lifetime parameters. The loss function for E to be trained is defined as follows,

$$E_{\text{loss}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

Where $y_i, \hat{y}_i$ represent the predicted and the ground-truth lifetime parameters, respectively.

Network training schedule. In this work, we employed three-stage training: generative model training stage, estimative model training stage, and *flimGANE* combination training stage. In generative model training stage, total iteration was set to be 2,000. Within each iteration, we randomly selected ~3% training samples from the pool of dataset. The discriminative model was updated five times while the generative model was kept untrainable, then the generative model is updated once while keeping the other one untrainable. In estimative model training stage, total iteration was set to be 5000. The aforementioned two training stages can be trained simultaneously and independently. Last, in *flimGANE* combination training stage, we combined the generative model and estimative model together, and fed *flimGANE* with randomly selected 10% training samples to update estimative model only for 100 times in each iteration, where the total iteration was set to be 500. Both the generative model and discriminative model were randomly initialized by Glorot uniform initializer. The generative model was optimized using Adam optimizer⁷ optimizer with a learning rate of 1×10^{-4} . The discriminative model was optimized using RMSprop optimizer with a starting learning rate of 5×10^{-5} . The final generative and discriminative models for each application in this work were selected at around 2000th iteration, which took 6.1h to train in generative model training stage. The estimator model which maps ground-truth fluorescence decay histogram to lifetime values was initialized by Glorot uniform initializer and optimized using Adam optimizer⁷ as well with learning rate of 1×10^{-3} . All the estimative models for specific applications were pre-trained for 5000 iterations, which took ~6 minutes to train in estimative model training stage. The generative models integrated with estimative models were then trained for 500 iterations, which took ~42 minutes to train in *flimGANE* combination training stage. Training without the discriminative loss and predictive cost can result in over-smoothed images, as the generative model optimizes only a specific group of statistical metrics. Therefore, it is imperative to incorporate discriminator to train the generative model well. A step-by-step training instruction and guideline, with several critical steps discussed and emphasized, are provided in **Supplementary Fig. 2**.

Software and hardware used for *flimGANE* development. This framework was implemented with TensorFlow framework version 1.12.0 and Python version 3.6.5 in the Microsoft Windows 10 operating system. The training was performed on a custom-built desktop equipped with single GeForce RTX 2070 XC GPU cards (NVIDIA) and a Core i7-9700K CPU @ 3.6 GHz (Intel).

Generation of the simulated database (*in silico*) with Monte Carlo (MC) method. First, multiple sets of ground truth are determined based on the lifetimes (τ_1 and τ_2) and the fraction amplitude (α_1). For each ground truth, different photon counts (pcs) and number of duplicates (e.g., 100) are assigned to construct the training dataset. For every training sample, it was assigned a value of short lifetime (τ_1), long lifetime (τ_2), fraction amplitude of short lifetime species (α_1), and photon counts (pcs). The IRF is obtained by averaging across all the pixels of the calibration image taken at the beginning of the experiment. These parameters were employed to generate the probability mass function that describes the distribution of the photon arrival time via the equation (10):

$$P(t) = N \left(IRF(t) \otimes \left[\alpha_1 e^{-\frac{t}{\tau_1}} + (1 - \alpha_1) e^{-\frac{t}{\tau_2}} \right] \right) \quad (10)$$

$$N(f(t)) = \frac{f(t)}{\text{sum}(f(t))}$$

Given the probability mass function, we can perform Monte Carlo simulation method to extract specified number (photon counts, *pcs*) of samples. Those extracted samples were then used to generate the simulated (degraded) decay histogram.

Fluorescence lifetime imaging microscopy (FLIM). The FLIM system were equipped with diode lasers providing 370 nm, 405 nm, 488 nm, and 635 nm lines and a supercontinuum white laser (SuperK EVO, NKT Photonics) providing a wavelength within 400 – 1000 nm. All the laser lines combined in Alba v5 passed a multi-band dichroic mirror to excite samples. Fluorescence emission light came back along the same path to the multi-band dichroic mirror, and went into 3 detectors, where each detector has its own pinhole. A Nikon inverted microscope with 60x, NA 1.2 water objective is used. An ASI XY automatic stage with motorized Z control is equipped the current setup. The photon counts were acquired with fastFLIM unit to build up the phase histogram. Here the laser repetition period we used is 50 ns, and it is divided into 256 bins.

Correction of the temporal shift between instrument response function (IRF) and acquired fluorescence decay histogram. Temporal shift occurred when the measurement is biased toward shorter or longer arrival time due to the noise. To characterize the temporal shift, the most common method is to perform model fitting with time shift as the additional parameter. During the process of curve fitting, the algorithms take temporal shift into consideration to find the optimal value for both lifetime parameters and time shift. However, this could take a large amount of time to obtain the optimal time shift parameter. Here we introduced a new analysis, Center of Mass Evaluation (CoME), to quantify the time shift parameter without the time-consuming curve fitting process. The mass center of the histogram, also known as the expected value of the histogram, was determined by the following equation,

$$CoM = \frac{\sum x_i h(x_i)}{\sum h(x_i)} \quad (11)$$

where $h(x_i)$ is the fluorescence decay histogram acquired from the experiment at the corresponding bin x_i .

We hypothesized that for certain lifetime, the difference of CoM between IRF and the fluorescence decay histogram should be the same. To verify the above hypothesis, we generated 250 simulated decay histograms under 150, 500, 1500, 5000 photon counts for various fluorescence lifetime values and calculated the mean and the standard deviation of the difference of CoM. **Supplementary Fig. 8** shows the calibration line for the distance of CoM between IRF and fluorescence decay histogram. Such a calibration line is employed to determine the amount of the temporal shift.

Implementation of center of mass evaluation (CoME) and trained *flimGANE*. To eliminate the “temporal shift” effect that may cause inaccurate lifetime estimation, this parameter was obtained by “center of mass” analysis with the control experiment and then employed to calibrate the measurement for each pixel. With IRF and the calibrated

fluorescence decay histogram as model inputs, *flimGANE* outputs two lifetime components and its relative ratio for each pixel, forming the FLIM image accurately and rapidly.

Fluorescence lifetime fitting with time-domain least-squares estimation (TD_LSE). Assume the fluorescence decay histogram contains either only one or two species:

$$f_{one\ species}(t, \alpha_1 = 1, \tau_1, \tau_2 = 0) = f(0) e^{-t/\tau_1} \quad (12)$$

$$f_{two\ species}(t, \alpha_1, \tau_1, \tau_2) = f(0) \left(\alpha_1 e^{-\frac{t}{\tau_1}} + (1 - \alpha_1) e^{-\frac{t}{\tau_2}} \right) \quad (13)$$

The expected measurement is the convolution of the fluorescence decay histogram and the IRF,

$$m(t, \alpha_1, \tau_1, \tau_2) = IRF(t) \otimes f(t, \alpha_1, \tau_1, \tau_2) \quad (14)$$

where τ_1 represents short lifetime, τ_2 represents long lifetime, α_1 is fraction amplitude of short lifetime species, and *IRF* is obtained from the experiments. The best fit parameters are determined by minimizing the sum of squared error

$$R^2 = \sum [y_i - m(t_i, A, \tau_1, \tau_2)]^2 \quad (15)$$

where y is the data to be fitted.

Fluorescence lifetime fitting with time-domain maximum likelihood estimation (TD_MLE). The observed fluorescence decay histogram is given by equation (14), where *IRF* represents the instrument response function, τ_1, τ_2 are the shorter and longer fluorescence lifetime, and α_1 is the fraction amplitude of shorter lifetime. The goal of MLE is to find the parameters that maximize the likelihood function, which is equivalent to minimizing the negative likelihood function. For the case of fluorescence lifetime analysis, we aim to minimize the negative Poisson log-likelihood function ($-L$), which is defined as follows,

$$-L = \sum_{i=0}^{N-1} (m_i - n_i \log m_i + \log n_i!) \quad (16)$$

where, n_i represents the number of the detected photons in i^{th} bin and m_i represents the expected measurement in i^{th} bin.

Fluorescence lifetime fitting with digital frequency-domain least-squares estimation (DFD_LSE). Due to the finite response of a system, the measured decay signal $f'(t)$ is the convolution of the fluorescence decay histogram and the IRF plus the noise, $n(t)$,

$$f'(t) = IRF \otimes f(t) + n(t) = IRF \otimes \left\{ f(0) \sum_{i=1}^N \alpha_i e^{-t/\tau_i} \right\} + n(t) \quad (17)$$

Thus, accurate FLIM data analysis typically requires the calibration of the IRF. In time-domain FLIM, the IRF can be measured by recording the scattered excitation light when using one-photon excitation, or the second-harmonic generation (SHG) signals for two-photon excitation by using a sample that yields strong SHG, such as urea crystal. In DFD-FLIM, the IRF is typically calibrated with a known fluorescence lifetime standard.

In TCSPC FLIM, a decay histogram $f'(t)$ is recorded at each pixel location. To present the data in frequency-domain, we take the Fourier Transform of the fluorescence decay histogram to get real and imaginary components.

$$\vec{F}(\omega) = \mathcal{F}(f'(t)) = \mathcal{F}(IRF \otimes f(t) + n(t)) = \mathcal{F}(IRF) * \mathcal{F}(f(t)) + \mathcal{F}(n(t)) \quad (18)$$

Notice that in frequency-domain, the convolution with the IRF is a simple multiplication of the IRF vector, and the noise is a simple addition of the noise vector. During the calibration procedure, we simply subtract the noise vector and divide the denoised vector by the IRF vector to get the emission fluorescent vector. If the emission fluorescent vector is from the standard calibration single component dye, the phasor will be on the semi-circle after the calibration. The later measurement of the unknown samples will be always noisy and IRF-free for finding the unknown fluorescent lifetime(s). The Fourier Transform of the fluorescence decay histogram can be separated into real number, $g(\omega)$, and imaginary number, $s(\omega)$.

$$g(\omega) = \frac{\int_0^\infty f'(t)\cos(\omega t) dt}{\int_0^\infty f'(t) dt} \quad (19)$$

$$s(\omega) = \frac{\int_0^\infty f'(t)\sin(\omega t) dt}{\int_0^\infty f'(t) dt} \quad (20)$$

Therefore, after calibration, each fluorescent decay histogram (without IRF and noise) of each pixel can be plotted as a single point, termed phasor, in the phasor plot by applying the sine and cosine transforms to the measured decay data, where the modulation frequency ω is the laser repetition angular frequency and is calculated by multiplying the laser repetition rate with 2π .

By solving the integral, we can then derive the following relationships between the phasor and the lifetime,

$$g(\omega) = \sum_{i=1}^N \frac{f_i}{1+\omega^2\tau_i^2} \quad (21)$$

$$s(\omega) = \sum_{i=1}^N \frac{f_i\omega\tau_i}{1+\omega^2\tau_i^2} \quad (22)$$

where N is the number of the fluorescent species, and f_i is the fractional contribution of the i^{th} species with the fluorescence lifetime τ_i .

Reduced chi-square (χ^2) for time-domain and digital frequency-domain data. In time-domain fitting process, the goodness-of-fit is evaluated by the reduced χ^2 , which is defined as

$$\chi^2 = \frac{1}{n-p-1} \sum_{k=1}^n \frac{[N(t_k) - N_c(t_k)]^2}{N_c(t_k)} \quad (23)$$

where n is the number of acquisition bin, p is the number of the fitting parameters, $N(t_k)$ is the number of photons collected at bin k , and $N_c(t_k)$ is the number of photons at bin k calculated from the fitting parameters and the decay model.

The reduced χ^2 for digital frequency-domain fitting is defined as

$$\chi^2 = \frac{1}{2n-p-1} \left\{ \sum_{k=1}^n \left[\frac{\varphi_k - \varphi_{ck}}{\sigma_{\varphi k}} \right]^2 + \sum_{k=1}^n \left[\frac{m_k - m_{ck}}{\sigma_{mk}} \right]^2 \right\} \quad (24)$$

where n is the number of modulation frequencies, p is the number of the fitting parameters, φ_k and m_k are the measured phase shift and modulation ratio at frequency k , φ_{ck} and m_{ck} are the calculated phase shift and modulation ratio at frequency k using the fitting parameters and the selected model, $\sigma_{\varphi k}$ and σ_{mk} are the uncertainties in the phase

shift and modulation ratio values at frequency k , respectively.

Pre-processing of FLIM data. Calibrating the phasor plot is the only procedure for digital frequency-domain fitting. The DFD-FLIM data measurements at each pixel location are composed of both the phase delay (φ) and the amplitude modulation ratio (m). The DFD-FLIM data at each pixel can be mapped to a single point called “phasor” in the phasor plot through a transform defined below, where ω is the modulation frequency, $g(\omega)$ and $s(\omega)$ represent the values at the two coordinates ($g(\omega)$, $s(\omega)$) of the phasor plot.

$$g(\omega) = m\cos(\varphi) \quad (25)$$

$$s(\omega) = m\sin(\varphi) \quad (26)$$

In order to establish the correct scale for the phasor analysis, the coordinates of the phasor plot need to be calibrated using a standard sample of known lifetime. This will include the calibration of the IRF and the background noise. In DFD-FLIM, this is done during experimental calibration prior to the data acquisition. There is no need to measure the IRF explicitly. The calibration procedure subtracts the noise and divides denoised data by the IRF to reveal the true fluorescent emission component(s). In time-domain FLIM, a directly recorded IRF data for the zero lifetime (scatter) will be measured. The IRF will then be used for lifetime fitting, with a convolution involved, and model analysis.

Post-processing of FLIM (intensity) image. Either a median or Gaussian filter can be applied before analyzing the data in phasor plot. The degree can be chosen from 0 to 10. When degree is 0, no filter is used. When the degree is 1, it runs through all the pixels once with a 3x3 filter. When degree is 2, it runs through all the pixels twice with a 3x3 filter.

For lifetime fitting, we can bin the data of each pixel to reduce the uncertainty. If the bin number is 0, it is the single pixel histogram or phase and modulation. If the bin number is 1, it borrows left, right, up and down 1 pixel, so total is 3x3 pixels as the selected pixel is at the center. In this work, the bin number is set to be 0.

Calibration of a FLIM system. Before imaging biological samples, it is necessary to calibrate the FLIM system using a fluorescence lifetime standard. The fluorescence lifetime for many fluorophores has been established under standard conditions, and any of these probes can be used for the calibration of the FLIM system. Since the fluorescence lifetime of a fluorophore is sensitive to its environment, it is critical to prepare the standards according to the conditions specified in the literature, including the solvent and the pH. It is also important to choose a standard fluorophore with excitation, emission, and fluorescence lifetime properties that are similar to those of the fluorophore used in the biological samples. For example, the dye, Coumarin 6, dissolved in ethanol (peak excitation and emission of 460 and 505 nm, respectively), with a reference lifetime of ~ 2.5 ns, is often used as the calibration standard for the CFPs. It is important to note that if the excitation wavelength is changed, it is necessary to recalibrate with another appropriate lifetime standard.

Preparation of fluorescence lifetime barcode beads and FLIM setup. The oligonucleotide-coated microbeads preparation was carried out using a modified protocol recommended by Bangs Laboratories. Briefly, 2 μ L (10 mg/mL) streptavidin-coated microbeads (CP01005, Bangs Laboratories Inc) were transferred into a 1.5 mL

centrifuge tube. The microbeads were washed with 20 μ L 1X PBS twice by centrifuging at 10K rpm for 3 min and resuspended in 1X PBS. The different ratios of mixed biotinylated single-strand DNA probes (**Probe1**: 5'Atto633-TGGTCGTGGGGCAACTGGGTT-biotin (3.5 ns) and **Probe2**: 5'Cy5-TTTTTTTTTTTTTT-biotin (1.9 ns) were purchased from Integrated DNA Technologies) were added and incubated for 15 min at room temperature with gentle mixing. The coated microbeads were then separated by centrifuging at 10K rpm for 3 min. The unbound biotinylated probes were removed by washing three times in 1X PBS. Then, two species coated beads are now ready for downstream applications. Here we demonstrated three different barcode beads (**Supplementary Table 8**) imaged by the FLIM. The FLIM images for fluorescence lifetime barcode beads were taken by the laser light focused through a 60X NA=1.2, water immersion objective (CFI Plan Apochromat Lambda 60XC, Nikon). A diode laser was used as an excitation source at 635 nm. The fluorescence was detected with an avalanche photodiode (SPCM-AQR-15, Perkin Elmer) after passing through a bandpass filter (685/40 nm, Semrock). FLIM images (512 $\times$ 512 pixels) were scanned three times with dwell time 0.04 ms/pixel. Cy5 in water (1 ns) was used for calibrating FLIM system.

Convallaria cover slide and FLIM setup. The convallaria (lily of the valley) cover slide was stained on 26mm x 76mm glass slides (RE-Cal01, Medical and Science Media). A supercontinuum white laser was used as an excitation source at 630/38 nm. The fluorescence was detected with an avalanche photodiode (SPCM-AQR-15, Perkin Elmer) after passing through a bandpass filter (685/40 nm, Semrock). The FLIM images were taken by the laser light focused through a 60X NA=1.2, water immersion objective (CFI Plan Apochromat Lambda 60XC, Nikon). FLIM images (512 $\times$ 512 pixels) were scanned one time (for low-photon-count condition) and three times (for medium-photon-count condition) with dwell time 0.1 ms/pixel. Cy5 in water (1 ns) was used for calibrating FLIM system.

Preparation of staining of live cells and FLIM setup. Live HeLa cells were seeded onto optical imaging 8-well Lab-Tek chambered cover glass (154534, Thermo Fisher Scientific) with cell density 70–90% confluent per well and grown overnight at 37°C in a humidified atmosphere with 5% CO₂ prior to staining. Cells were maintained in DMEM/F12 medium (11320082, Thermo Fisher Scientific) supplemented with 10% heat-inactivated fetal bovine serum (16140071, Thermo Fisher Scientific) and 50 U/mL penicillin-streptomycin (15070063, Thermo Fisher Scientific). CellMaskTM Green (C37608, Thermo Fisher Scientific) or CellMaskTM Red (C10046, Thermo Fisher Scientific) plasma membrane stain (1 μ g/mL) were used to stain the plasma membrane of live cells for 10 min at 37°C. The staining solution was removed from the chambered cover glass. Then, the live cells were washed with PBS three times. The nucleus was stained by the permeable Hoechst 33342 dye (62249, Thermo Fisher Scientific) for 10 min at 37°C and washed with PBS three times. Cells were then kept in the phenol red-free DMEM/F12 (21041025, Thermo Fisher Scientific) for the FLIM images acquisition. A diode laser was used as an excitation source at 405, 488 and 635 nm. The FLIM images were taken by the laser light focused through a 60X NA=1.2, water immersion objective (CFI Plan Apochromat Lambda 60XC, Nikon). The fluorescence was detected with an avalanche photodiode (SPCM-AQR-15, Perkin Elmer) after passing through a bandpass filter (494/34, 531/40, 685/40 nm, Semrock). FLIM images (512 $\times$ 512 pixels) were taken with dwell time 0.1 and 0.2 ms/pixel, respectively.

Alexa 405 in water (3.6 ns), rhodamine 110 in water (4 ns) and Cy5 in water (1 ns) was used for calibrating FLIM system.

Preparation of intramolecular FRET glucose sensor. Triple negative breast tumor cell line, MDA-MB-231, was obtained from the American Type Culture Collection (ATCC) and grown in high-glucose (25 mM) DMEM/F12 culture medium (11320082, Thermo Fisher Scientific) containing 10% heat-inactivated fetal bovine serum (16140071, Thermo Fisher Scientific) and 50 U/mL penicillin-streptomycin (15070063, Thermo Fisher Scientific). The plasmid carrying the glucose FRET sensor, pcDNA3.1 FLII12Pglu-700uDelta6 (a gift from Wolf Frommer), was obtained from Addgene. Prior to transfection, MDA-MB-231 cells were seeded in a 6-well plate and allowed with cell density 70–90% confluent per well. Transfections were performed using Lipofectamine™ LTX and Plus™ reagent (15338100, Thermo Fisher Scientific) according to manufacturer's instructions. Transfection medium, Opti-MEM™ I Reduced Serum Medium (11058021, Thermo Fisher Scientific), contained no serum or antibiotics. Six hours post-transfection, the medium was replaced with DMEM culture medium. Three days post-transfection, medium was replaced with DMEM containing 100 µg/mL G418 (A1720, Sigma-Aldrich) for selection. After two weeks of selection, the cells were sorted by flow cytometry (FACSaria, BD Biosciences) based on YFP expression. MDA-MB-231 cells transfected with the FRET glucose sensor were seeded onto optical imaging 8-well Lab-Tek chambered cover glass (154534, Thermo Fisher Scientific) with cell density 70–90% confluent per well and grown overnight at 37°C in a humidified atmosphere with 5% CO₂. The medium was replaced with glucose-free DMEM culture medium for 24 hours before FLIM image acquisition. The FLIM images were taken by the laser light focused through a 20X objective (UPLSAPO, Olympus). The fluorescence of CFP and YFP were detected by two avalanche photodiodes (SPCM-AQR-15, Perkin Elmer) after passing through the bandpass filters (445/40, 585/40 nm, Semrock), respectively. FLIM images (256×256 pixels) were scanned three times with dwell time 0.1 ms/pixel. Alexa 405 in water (3.6 ns) was used for calibrating FLIM system.

Autofluorescence FLIM for cellular metabolism and FLIM setup. Live HeLa cells were seeded onto optical imaging 8-well Lab-Tek chambered cover glass (154534, Thermo Fisher Scientific) with cell density 70–90% confluent per well and grown overnight at 37°C in a humidified atmosphere with 5% CO₂. Before taking an autofluorescence FLIM image, the medium was replaced with phenol red-free complete medium. A diode laser was used as an excitation source at 405 nm. The FLIM images were taken by the laser light focused through a 60X NA=1.2, water immersion objective (CFI Plan Apochromat Lambda 60XC, Nikon). The autofluorescence of NAD(P)H /FAD and were detected by two avalanche photodiodes (SPCM-AQR-15, Perkin Elmer) after passing through the bandpass filters (445/40, 531/40 nm, Semrock), respectively. FLIM images (512×512 pixels) were scanned one time with dwell time 0.1 ms/pixel. Alexa 405 in water (3.6 ns) was used for calibrating FLIM system.

Intensity contrast segmentation for mitochondria in Autofluorescence FLIM. Intensity contrasts of both FAD and NAD(P)H images were normalized to the scale, 0 – 1. Normalized value for each pixel is determined by the following equations:

$$\tau_{normalized} = \frac{\tau_{original} - \tau_{min}}{\tau_{max} - \tau_{min}} \quad (27)$$

Given both normalized images, the locations of segmented mitochondrial were verified where the value presenting the specific location was both greater than the threshold. In this work, the threshold was set to be 0.25. The locations of segmented cytoplasm were determined between 0.16 and 0.25. The locations of segmented nuclei were determined between 0.06 and 0.16.

Quality index (QI) quantification for simulated decay histogram. To calculate the QI for the fluorescence decay histogram, we assume the amounts of signals is the total photon counts, and the amount of noise is the deviation of fluorescence decay histogram from the ground-truth decay histogram. The relationship can be described as the following equations:

$$QI = \frac{S}{\sum_i |y_i - r_i|} \quad (28)$$

where S represents the total photons of the fluorescence decay histogram, y_i is the fluorescence decay histogram on i^{th} bin, and r_i represents the ground-truth fluorescence decay histogram on i^{th} bin.

Structural similarity index (SSIM)⁸. SSIM is a method used for measuring the similarity between two images. First, two images were smoothed with Gaussian filter ($\sigma = 1.5$ pixels). Then SSIM index was obtained for each window of a pair of sub-images. It was calculated for squared windows, centered at the same pixel (dx, dy) of two images. The length of a side of the square window was eleven pixels. The SSIM (x, y) was then obtained for two windows in two images (x , and y) as follows,

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (29)$$

where, μ_x, μ_y are the average of pixel intensity of window x and y , $\sigma_x, \sigma_y, \sigma_{xy}$ are standard deviation of window x and y , and covariance of two windows, c_1 , and c_2 are $(0.01 \times L)^2$ and $(0.03 \times L)^2$, respectively, and L equals to the data range of the lifetime image. Then $SSIM(x, y)$ will be averaged over all the area. The average of SSIM, $\overline{SSIM}$, will be outputted, serving as “SSIM” for each pair of images in this work.

Mean-squared errors (MSE) and peak signal-to-noise ratio (PSNR)⁹. PSNR is a method used for measuring the quality of the image comparing to the reference image. Higher PSNR (higher score) represents better image quality. Given the reference $n \times m$ FLIM image I and the reconstructed FLIM image K ,

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (30)$$

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (31)$$

where, MAX_I is the maximum possible lifetime value of the reference image I .

Visual information fidelity (VIF)¹⁰. VIF is a full-reference image quality model based on natural scene statistical (NSS) model and human visual system (HVS). In particular, VIF models the perceived picture quality as a communication channel, where the information of reference signal $I(C,E|s)$ and distorted signal $I(C,F|s)$ are first extracted by

$$I(\vec{C}^N; \vec{E}^N | s^N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{s_i^2 \lambda_k}{\sigma_n^2} \right) \quad (32)$$

$$I(\vec{C}^N; \vec{F}^N | s^N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{g_i^2 s_i^2 \lambda_k}{\sigma_v^2 + \sigma_n^2} \right) \quad (33)$$

where s_i represents the i^{th} positive scalars of the random filed (RF) in Gaussian Scale Mixture (GSM) model, C_u and λ_k are the covariance and the corresponding eigenvalues of the Gaussian vector RF, g_i is the i^{th} deterministic scalar gain field in the distortion model as well as the variance of the stationary additive Gaussian noise RF, C_v , and σ_n^2 represents the variance of the visual noise in the human visual system model. These parameters are estimated with maximum likelihood estimation.

The final VIF score is then defined as the ratio between two mutual information measures

$$VIF = \frac{\sum_{j \in channels} I(\vec{C}^{N,j}; \vec{F}^{N,j} | s^{N,j})}{\sum_{j \in channels} I(\vec{C}^{N,j}; \vec{E}^{N,j} | s^{N,j})} \quad (34)$$

where, $I(\vec{C}^{N,j}; \vec{E}^{N,j} | s^{N,j})$ and $I(\vec{C}^{N,j}; \vec{F}^{N,j} | s^{N,j})$ represent the information that can ideally be extracted from a particular channel, j , in the reference and test images respectively.

Apparent and average lifetime for mixture of fluorophores. We referred to technique note from ISS (http://www.iss.com/resources/pdf/technotes/FLIM_Using_Phaseor_Plots.pdf) to define the apparent lifetime and average lifetime^{11, 12}. Given the multi-exponential decay model defined as follows,

$$I(t) = \sum_{i=1}^n \alpha_i e^{(-t/\tau_i)} \quad (35)$$

where, τ_i represents the lifetime, α_i is the amplitude of the components at $t = 0$, also called pre-exponential factor, and n is the number of the components. Fractional contribution, f_i , of each decay histogram to the steady-state intensity is determined by

$$f_i = \frac{\alpha_i \tau_i}{\sum_j \alpha_j \tau_j} \quad (36)$$

Apparent lifetime is defined by,

$$\tau_{ap} = \sum_j \alpha_j \tau_j \quad (37)$$

Average lifetime is defined by,

$$\bar{\tau} = \frac{\alpha_1 \tau_1^2 + \alpha_2 \tau_2^2}{\alpha_1 \tau_1 + \alpha_2 \tau_2} = f_1 \tau_1 + f_2 \tau_2 \quad (38)$$

Discriminability test. Discriminability test was utilized to assess the “limits of lifetime differentiation.” This test was performed using a reference lifetime of 2.00 ns and under 5 photon-count conditions (**Supplementary Table**

2 and Supplementary Fig. 21a). 500 simulated decays were generated for each τ , with $\Delta\tau$ ranging from 0.01-0.30 ns (i.e., $\tau = 2.00, 2.01, \dots 2.30$ ns). The mean lifetime was extracted from 500 lifetime estimates (by Gaussian fitting of the histogram) and each $\Delta\tau$ was repeatedly tested for 20 times. Using the p -value < 0.05 in two-sided KS test as a criterion, if the p -value was less than 0.05 for 70% of the repeated tests, we determined that there was a significant difference between two lifetime distributions. For instance, 2.00 ns decay and 2.03 ns decays were indistinguishable as their p -value was 0.33 (**Supplementary Fig. 21b**). In contrast, 2.00 ns decay and 2.15 ns decay were distinguishable with a p -value of 1.95×10^{-16} (**Supplementary Fig. 21c**). We noticed in all 5 photon-count conditions, *flimGANE* and *TD_MLE* had the same discriminability (**Supplementary Fig. 21d**). As expected, higher photon counts per pixel improved the discriminability.

***flimGANE* outperforms *TD_MLE* at ultra-low-photon-count conditions.** To ensure a stable training process, in this work, Wasserstein GAN algorithm (WGAN⁴) was implemented, which changes the value function to be based on Wasserstein distance $W(p_r, p_g)$ (equation (39)):

$$W(p_r, p_g) = \inf_{\gamma \sim \Pi(p_r, p_g)} \mathbb{E}_{x, y \sim \gamma} \|x - y\| \quad (39)$$

where $\Pi(p_r, p_g)$ represents the set of all joint distributions $\gamma(x, y)$ whose marginals are p_r and p_g , respectively. In principle, W represents the lowest work to convert one distribution into another. WGAN can resolve the issue of vanishing gradient, thus eliminating the factor that could destabilize training⁴. Indeed, in our simulations, we have successfully trained the model to learn to reconstruct the ground-truth “mimicking” fluorescence decay histograms (**Supplementary Fig. 16**), which could not be achieved by the classical GAN.

Although both GAN and MLE can be regarded as an optimization problem, they are fundamentally different. MLE requires a likelihood function while GAN does not. Maximum likelihood approach is equivalent to minimizing the KL divergence (equations (40-43)). Assuming θ represents the unknown parameters to be estimated (e.g., α_1, τ_1, τ_2 in fluorescence lifetime analysis) and $\hat{\theta}$ is a set of values that maximizes the likelihood, determined by the estimated decay histogram $p_\theta(y)$ and the observed data, and match $p_\theta(y)$ with the ground-truth decay histogram $p_{\theta^*}(y)$:

$$\hat{\theta} = \operatorname{argmax}_{\theta} \prod_{i=1}^n p_\theta(y_i) = \operatorname{argmax}_{\theta} \sum_{i=1}^n \log p_\theta(y_i) = \operatorname{argmax}_{\theta} \frac{1}{n} \sum_{i=1}^n \log p_\theta(y_i) \quad (40)$$

When $n \rightarrow \infty$

$$\frac{1}{n} \sum_{i=1}^n \log p_\theta(y_i) \approx \mathbb{E}[\log p_\theta(y)] \quad (41)$$

Instead of maximizing a positive function, we can minimize the corresponding negative function, where $\hat{\theta}$ can be formulated as follows:

$$\hat{\theta} = \operatorname{argmax}_{\theta} \mathbb{E}[\log p_\theta(y)] = \operatorname{argmin}_{\theta} -\mathbb{E}[\log p_\theta(y)] \quad (42)$$

The KL divergence between $p_{\theta^*}(y)$ and $p_\theta(y)$ is:

$$\begin{aligned}
D_{KL}(p_{\theta^*}(y)||p_{\theta}(y)) &= \int_y p_{\theta^*}(y) \log \frac{p_{\theta^*}(y)}{p_{\theta}(y)} dy = \mathbb{E} \left[\log \frac{p_{\theta^*}(y)}{p_{\theta}(y)} \right] \\
&= \mathbb{E}[\log p_{\theta^*}(y) - \log p_{\theta}(y)] = \mathbb{E}[\log p_{\theta^*}(y)] - \mathbb{E}[\log p_{\theta}(y)]
\end{aligned} \tag{43}$$

The first term in equation (43) can be ignored when minimizing the KL divergence for θ . This results in the same formula in equation (42). It is clear that the KL divergence calculation determines the quality of MLE method. It was noted that KL divergence may not output a value of useful information under certain conditions, but Wasserstein distance can provide a smoother and usable gradient everywhere⁴. Please see the following examples.

We have two multivariate Gaussian distribution N_1, N_2 where these distributions have dimension k , means μ_i , and covariance matrices Σ_i , for $i = 1$ and 2 . The Wasserstein distance and the KL divergence between N_1 and N_2 are formulated as follows:

$$W_2(N_1, N_2) = \|\mu_1 - \mu_2\|_2^2 + \text{tr}(\Sigma_1 + \Sigma_2 - 2 \left(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2} \right)^{1/2}) \tag{44}$$

$$D_{KL}(N_1||N_2) = \frac{1}{2} (\text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) - k + \ln(\frac{\det \Sigma_2}{\det \Sigma_1})) \tag{45}$$

Considering $\Sigma_1 = \Sigma_2 = I_k$ and $\mu_1 \neq \mu_2$, the trace term in Wasserstein is 0 and the trace term in the KL divergence will be 0 when combined with the $-k$ term and the log-determinant ratio, so these two quantities become:

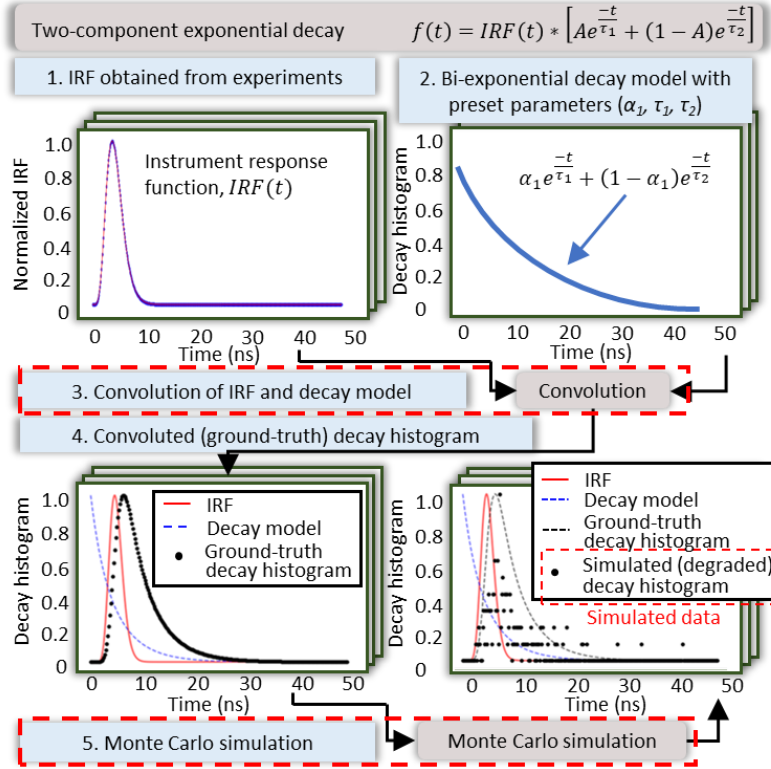
$$W_2(N_1, N_2) = \|\mu_1 - \mu_2\|_2^2 \tag{46}$$

$$D_{KL}(N_1||N_2) = +(\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) \tag{47}$$

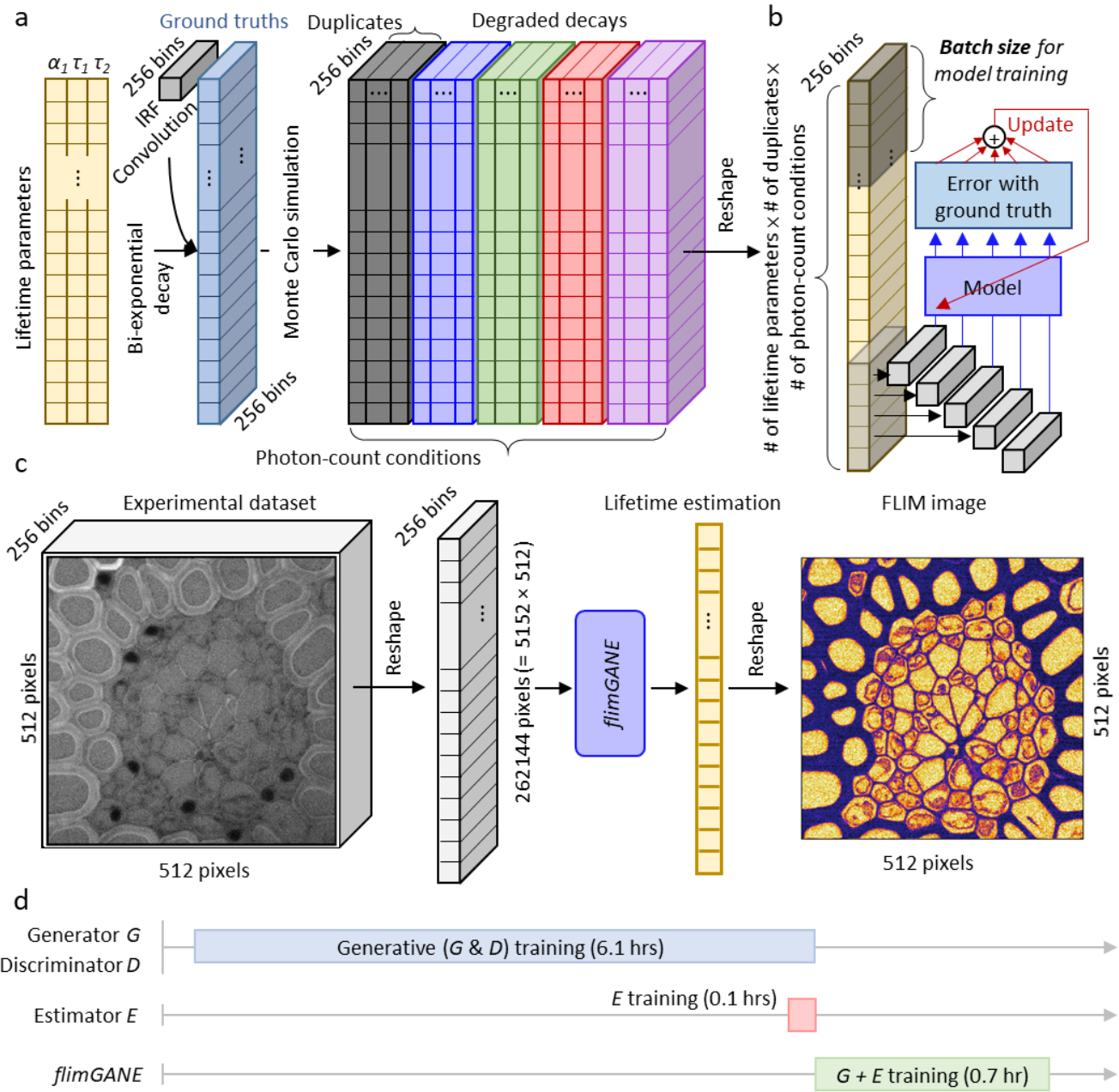
The Wasserstein distance does not change when the variance changes since the Wasserstein distance is a distance function in the joint support spaces of the two probability measures. In contrast, the KL divergence does change with the variance because the KL divergence is a divergence and this divergence changes based on the spread of the distributions. While these two measures (W and KL) have advantages and disadvantages in different applications^{17, 18, 19}, when handling the low-photon-count data, our simulations showed more reliable results from Wasserstein distance calculation-based training process, which led to the better performance of *flimGANE* in analyzing low-photon-count data.

Here we prepared simulated decay histograms (ranged from 1 to 5 ns) with different photon counts (50, 100, 500 and 1500 photons per pixel) and their corresponding “artificial” ground-truth decay histograms (G_{output}) were generated from *Generator*. The G_{output} lifetimes were determined by fitting with *TD_MLE* (termed $TD_MLE_{G_output}$). The accuracy of lifetime estimation with *TD_MLE*, $TD_MLE_{G_output}$ and *flimGANE* was evaluated in **Supplementary Fig. 17**. $TD_MLE_{G_output}$ produces results better than those of *TD_MLE* under ultra-low-photon-count condition, validating that the *Generator* in *flimGANE* can reliably reconstruct the “artificial” ground-truth decay histograms. As expected, all the lifetime determination methods generate more accurate results with more photon counts in the decay histogram.

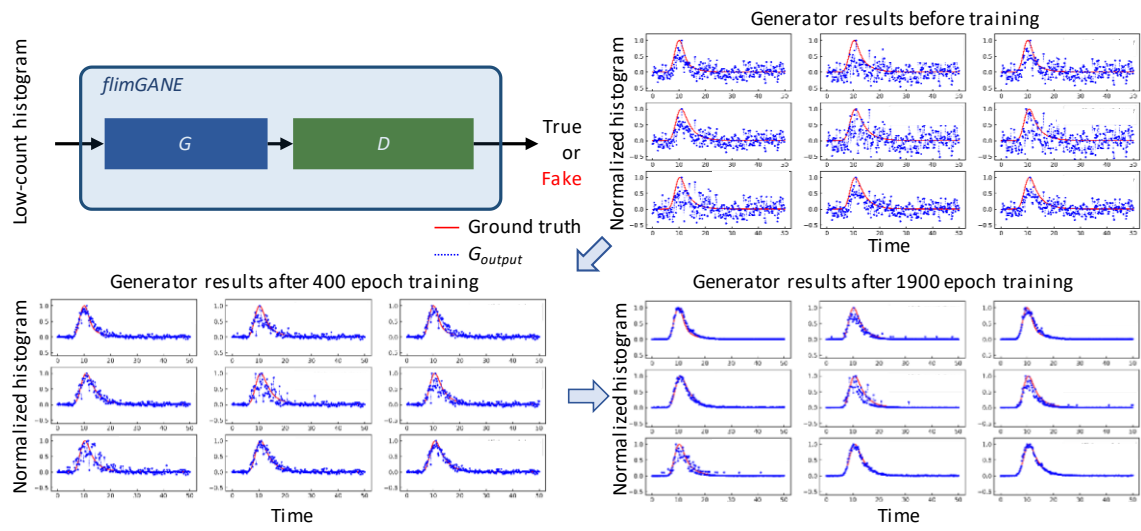
In conclusion, we believe KL divergence and Wasserstein distance are two measures which have different performance under low-photon-count conditions (**Supplementary Table 3-5**).



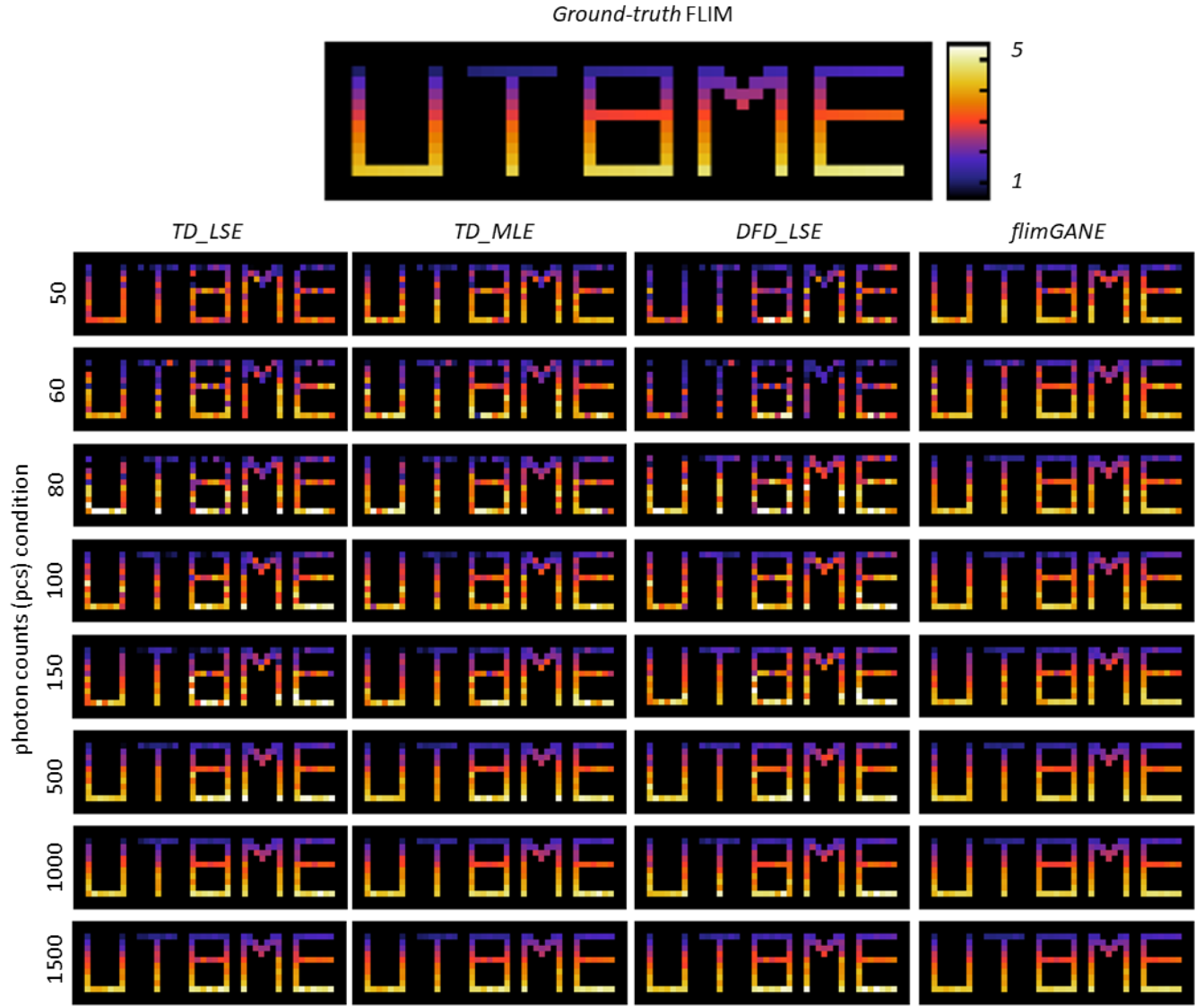
Supplementary Fig. 1 Generation of the simulated decay histogram with lifetime parameters and photon counts by MC simulation. First, instrument response function (IRF) was obtained from the experiments. Given the pre-defined lifetime parameters, a bi-exponential decay model is then obtained. After convolution with the experimental IRF, we normalized the convoluted decay histogram to have unit area under curve. Regarding the normalized function as probability mass function (pmf), which is served as the ground-truth decay histogram, we implemented MC simulation based on the given PMF to generate the samples with the size as the assigned photon counts. The simulated decay histogram was then obtained from the resulting samples plotted as histogram.



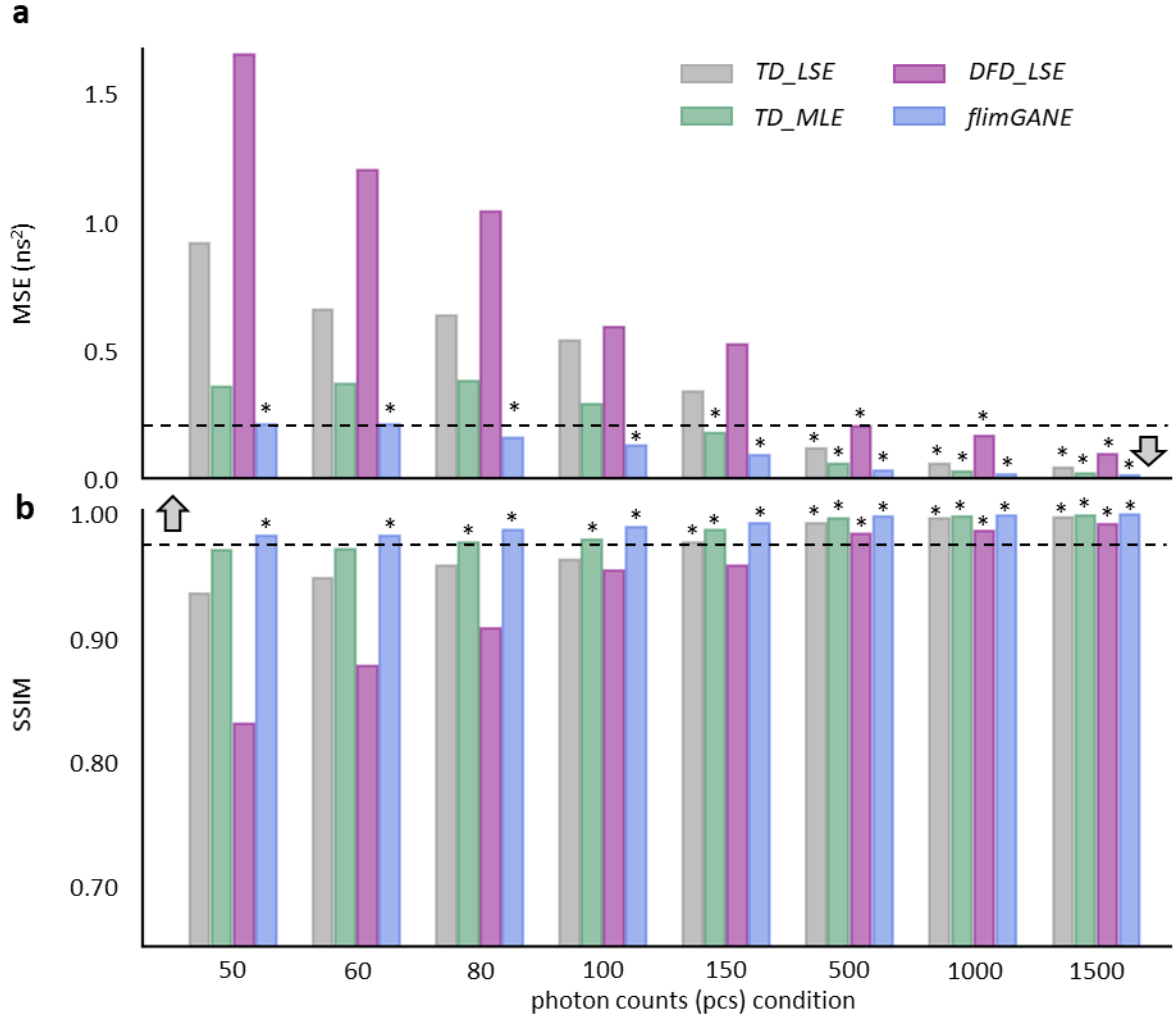
Supplementary Fig. 2 Schematics of simulated training dataset and training procedures for *flimGANE*. **a** Workflow of simulation process to generate the training dataset. **b** A subset of dataset (determined by the batch size, e.g., 5) was randomly selected for model training at each iteration. **c** Workflow of data processing steps to generate a FLIM image from an experimental dataset. **d** Timeline of *flimGANE* training process. The *G* training and *E* training processes can be performed in parallel. Then we combine the *G* and *E* together to conduct *flimGANE* training process, which is the last step before the real-world implementation.



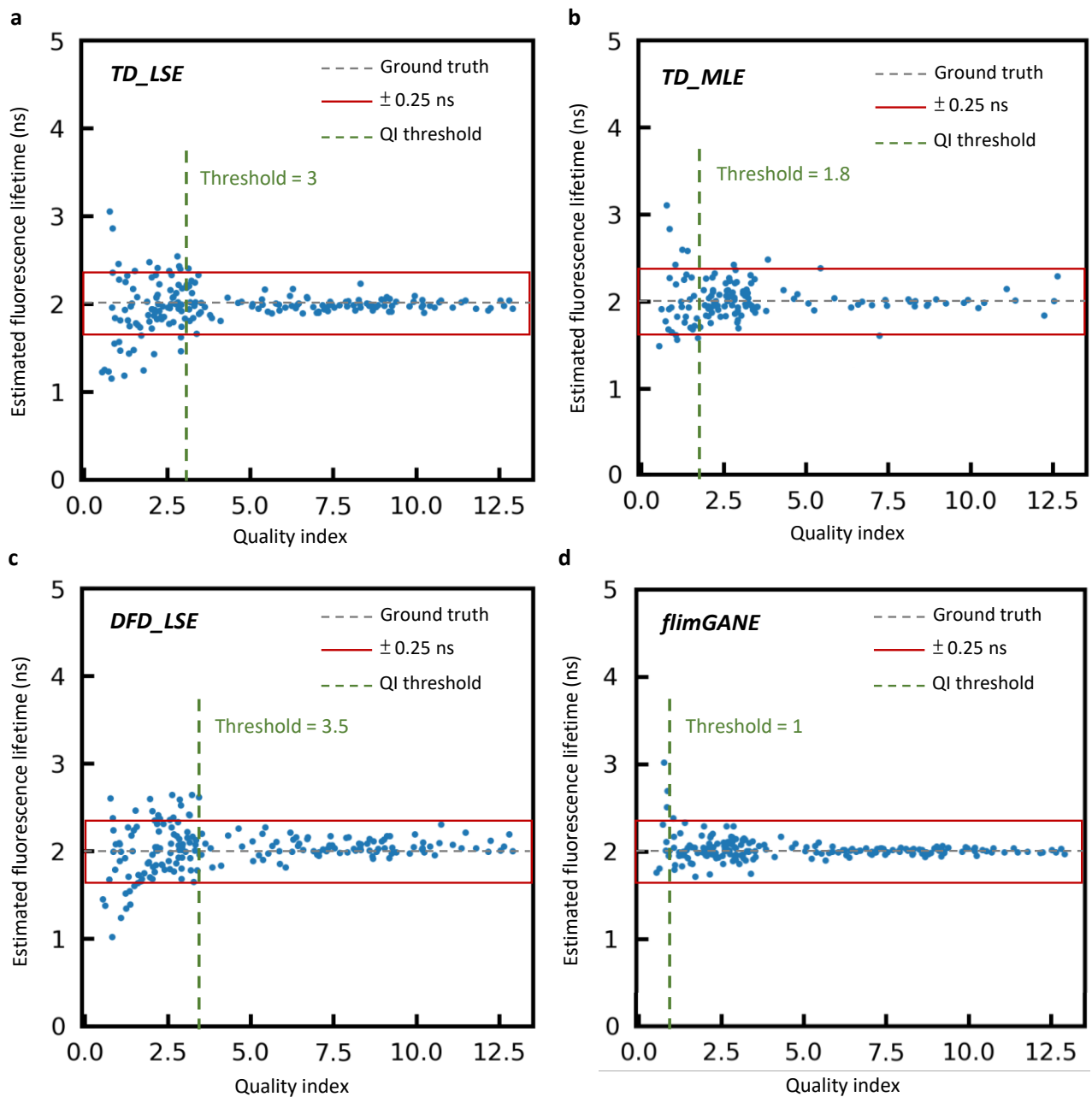
Supplementary Fig. 3 Generator (*G*), a subnet from *flimGANE*, can transform low-photon-count decay histogram into high-photon-count one. Given the *flimGANE* framework, the normalized low-photon-count decay histogram was transformed into the normalized ground-truth mimicking histogram. At the beginning of the training stage, the output from the generator was chaotic. The generator-inferred fluorescence decay histogram gradually matched with the ground truth.



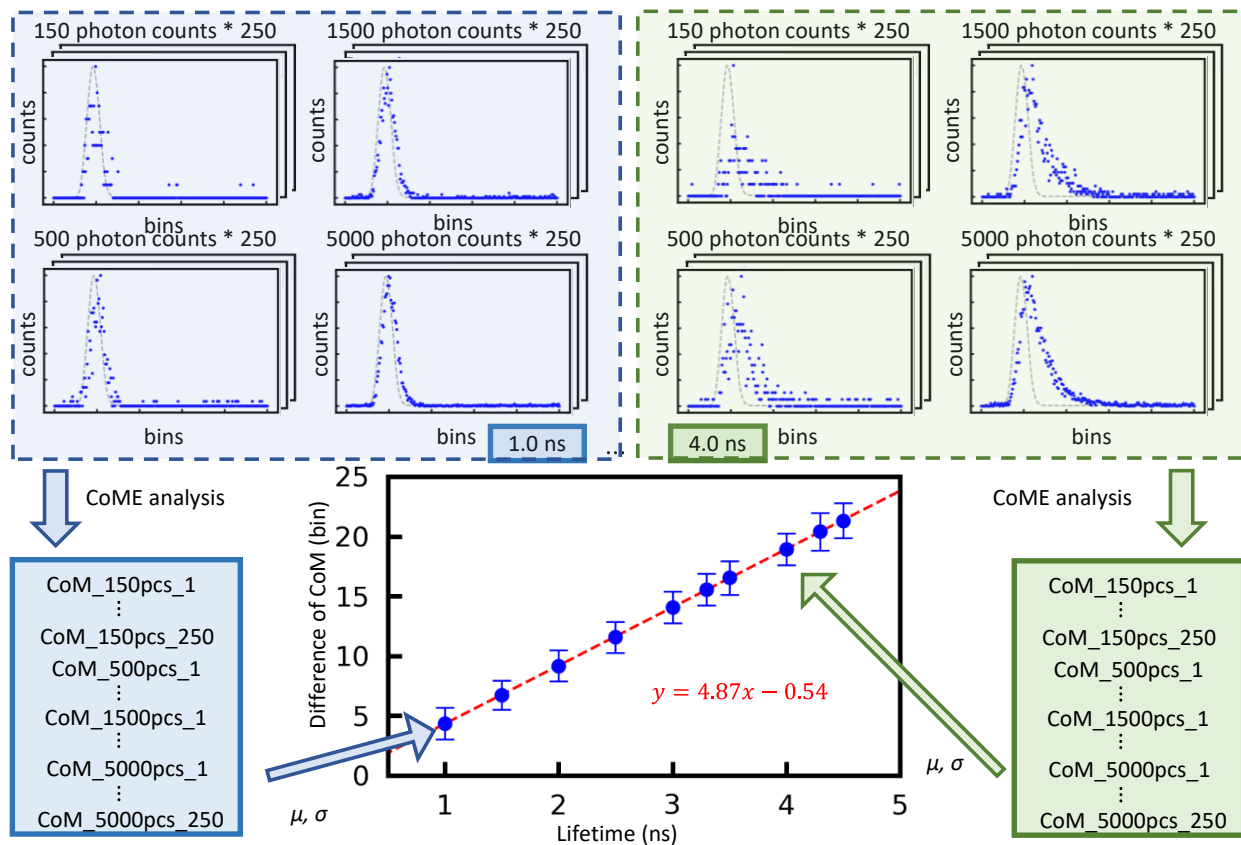
Supplementary Fig. 4 *flimGANE* outperformed other analysis methods in FLIM images reconstruction at various photon-count conditions *in-silico*. FLIM image was generated through MC simulation with different conditions (lifetime, $\tau = 1.0 \sim 5.0$ ns; photon counts, $pcs = 50, 60, 80, 100, 150, 500, 1000, 1500$). It was obvious that *flimGANE* outperformed other analysis methods. As expected, under extremely low-photo-count condition (< 80 counts), *TD_MLE*, *TD_LSE*, and *DFD_LSE* are unable to make accurate estimation. Under low-photon-count condition (100, 150 counts), *TD_MLE* was close to the accurate estimation; however, *TD_LSE* and *DFD_LSE* can only generate accurate FLIM image with high-photon-count conditions (> 500 counts).



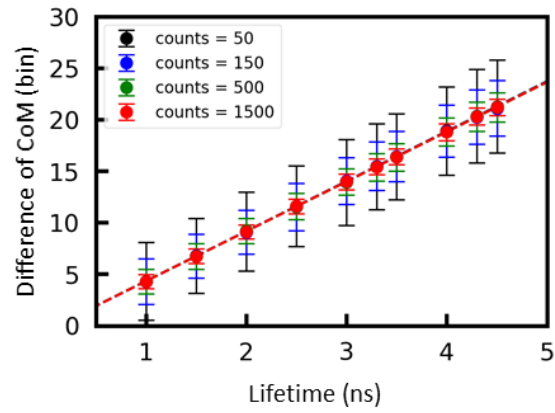
Supplementary Fig. 5 The performance of different analysis methods was evaluated by MSE and SSIM. The reconstructed “UT BME” FLIM images obtained by different analysis methods were evaluated by mean squared error (MSE, threshold = 0.21) and structural similarity index (SSIM, threshold = 0.98). **a-b** The performance of different analysis methods was evaluated under eight conditions, 50, 60, 80, 100, 150, 500, 1000, 1500 photon counts. Under ultra-low-photon-count condition (50-100 counts), the MSE and SSIM between the reconstructed FLIM images from *flimGANE* and the ground-truth FLIM images were less than 0.21 and larger than 0.98. *TD_MLE* provided MSE less than 0.21 and SSIM larger than 0.98 under low-photon-count condition. *TD_LSE* and *DFD_LSE* could provide MSE less than 0.2 as the photon counts were greater than 500.



Supplementary Fig. 6 The performance of different analysis methods was evaluated by quality index (QI). *flimGANE* outperforms other analysis methods in fluorescence lifetime estimation from low-photon-count (noisy) decay histogram. The fluorescence decay histograms were simulated with 2 ns as pre-defined lifetime and various photon-count conditions (100 ~ 10k counts), generating the simulated decay histogram with quality index (QI) ranging from 0.5 to 13.0. The tolerance of the lifetime estimate is 0.25 ns (red box). We can observe that *flimGANE* enables accurate lifetime estimation under poor QI condition as 1, which is much smaller than *TD_MLE* (1.8), *TD_LSE* (3), and *DFD_LSE* (3.5).

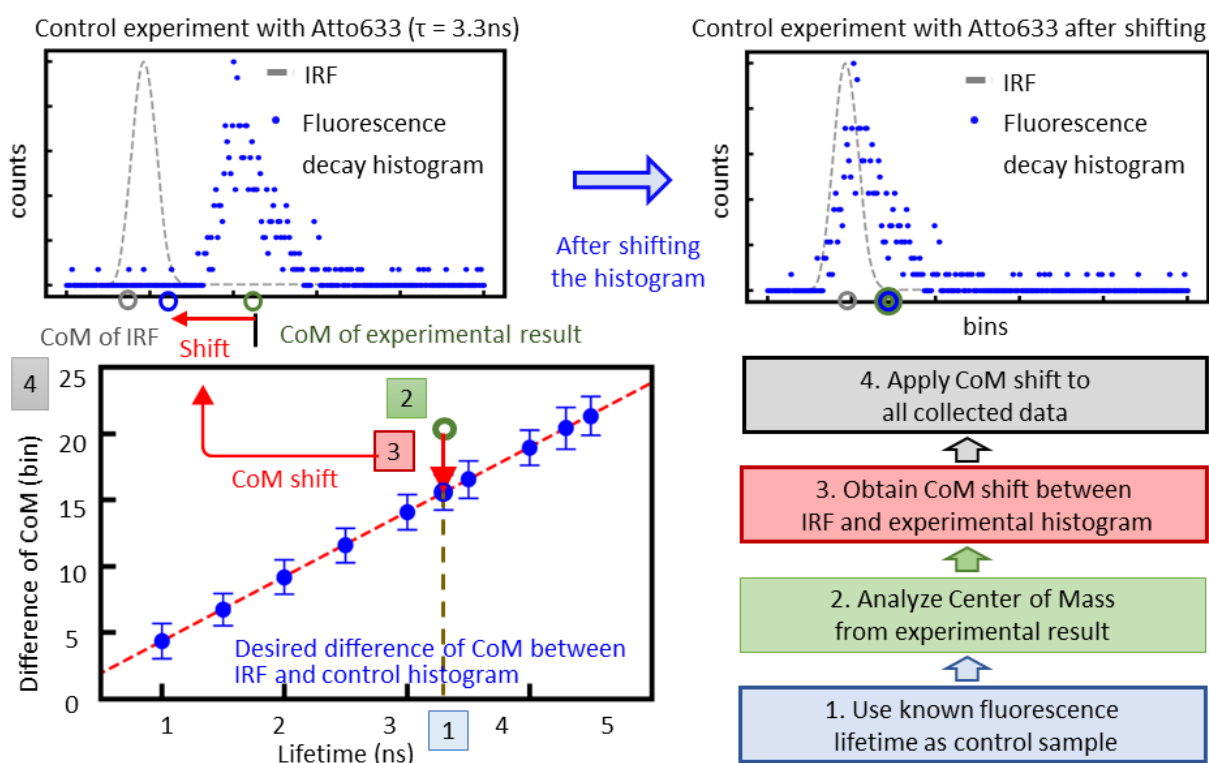


Supplementary Fig. 7 Difference of center of mass (CoM) between the IRF and decay histogram was demonstrated positively correlated. In order to calibrate the temporal shift between IRF and fluorescence decay histogram, we generated a calibration line to adjust the shift difference between them. First, we performed MC simulation to generate 1,000 simulated decay histograms for each lifetime value (1.0 ~ 4.5 ns) at different conditions (150, 500, 1500, 5000 photon counts). Second, we performed CoM analysis to obtain the mass center of each decay histogram and IRF. All the analysis results were plotted with the x axis as lifetime value and y axis as the distance of mass centers between fluorescence decay histogram and IRF (error bars, standard deviation, $n = 1000$). A linear regression line, calibration line, is obtained by fitting the analysis results.

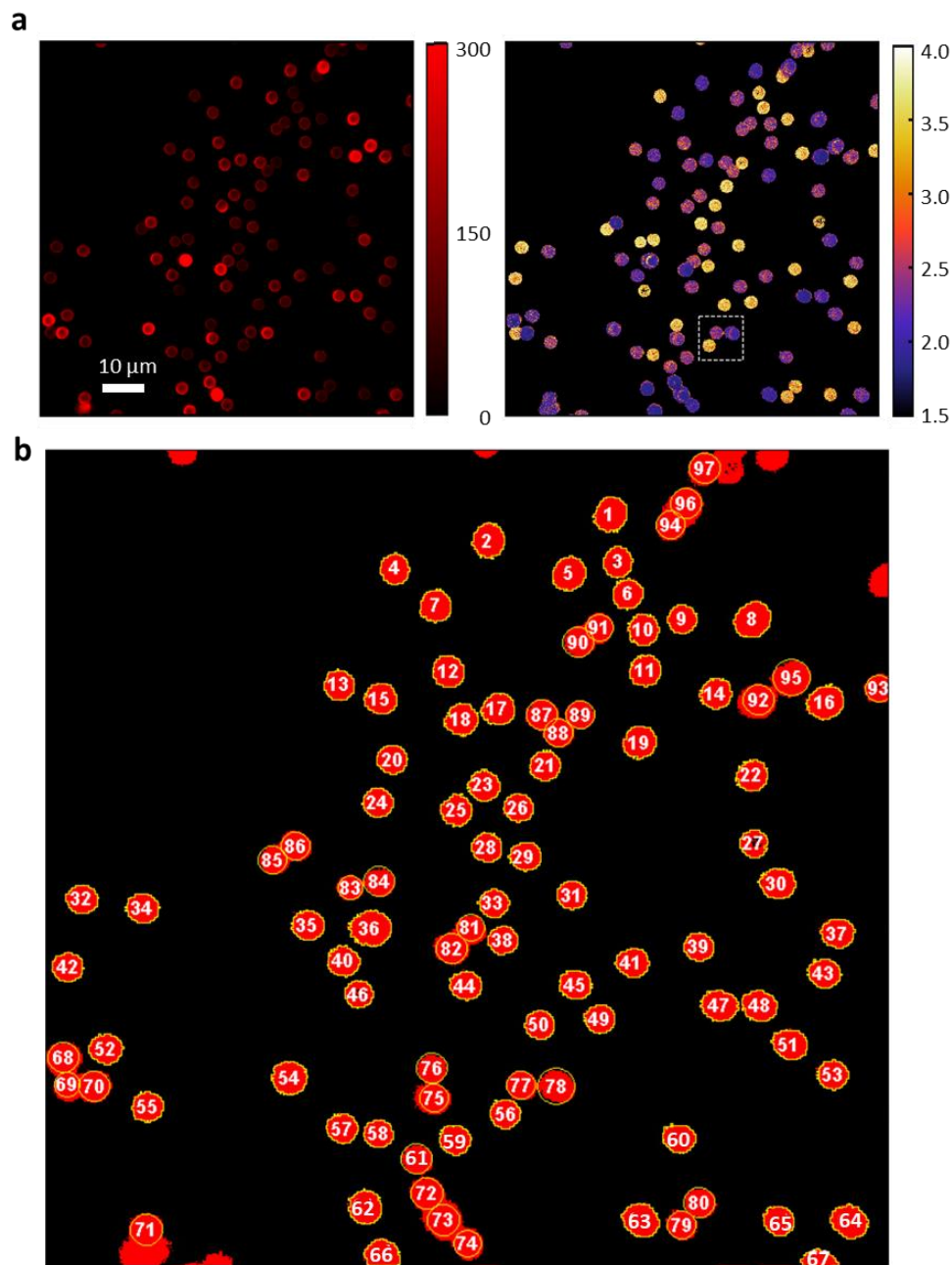


Supplementary Fig. 8 Calibration line for center of mass evaluation (CoME) was consistent over different photon counts.

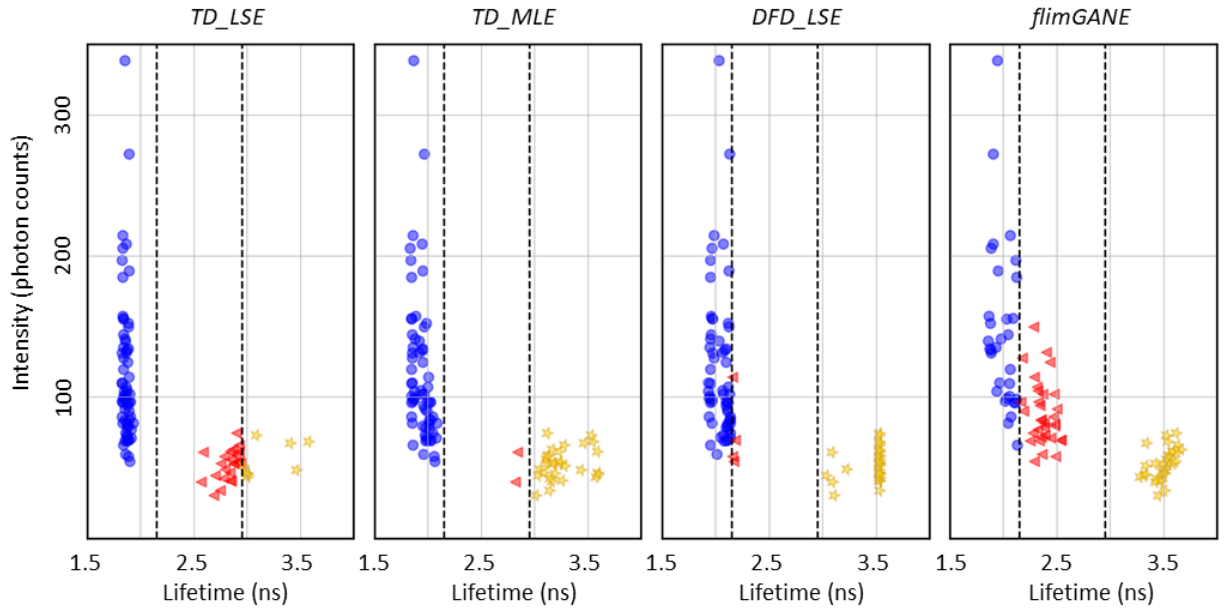
The analysis results in **Supplementary Fig. 7** were plotted with distinct photon-count conditions (error bars, standard deviation, $n = 250$). It was demonstrated that the calibration is independent of the photon counts. The mean of the difference of CoM between IRF and decay histograms was the same at varying conditions. Moreover, it is expected to observe a higher variance at low-photon-count condition.



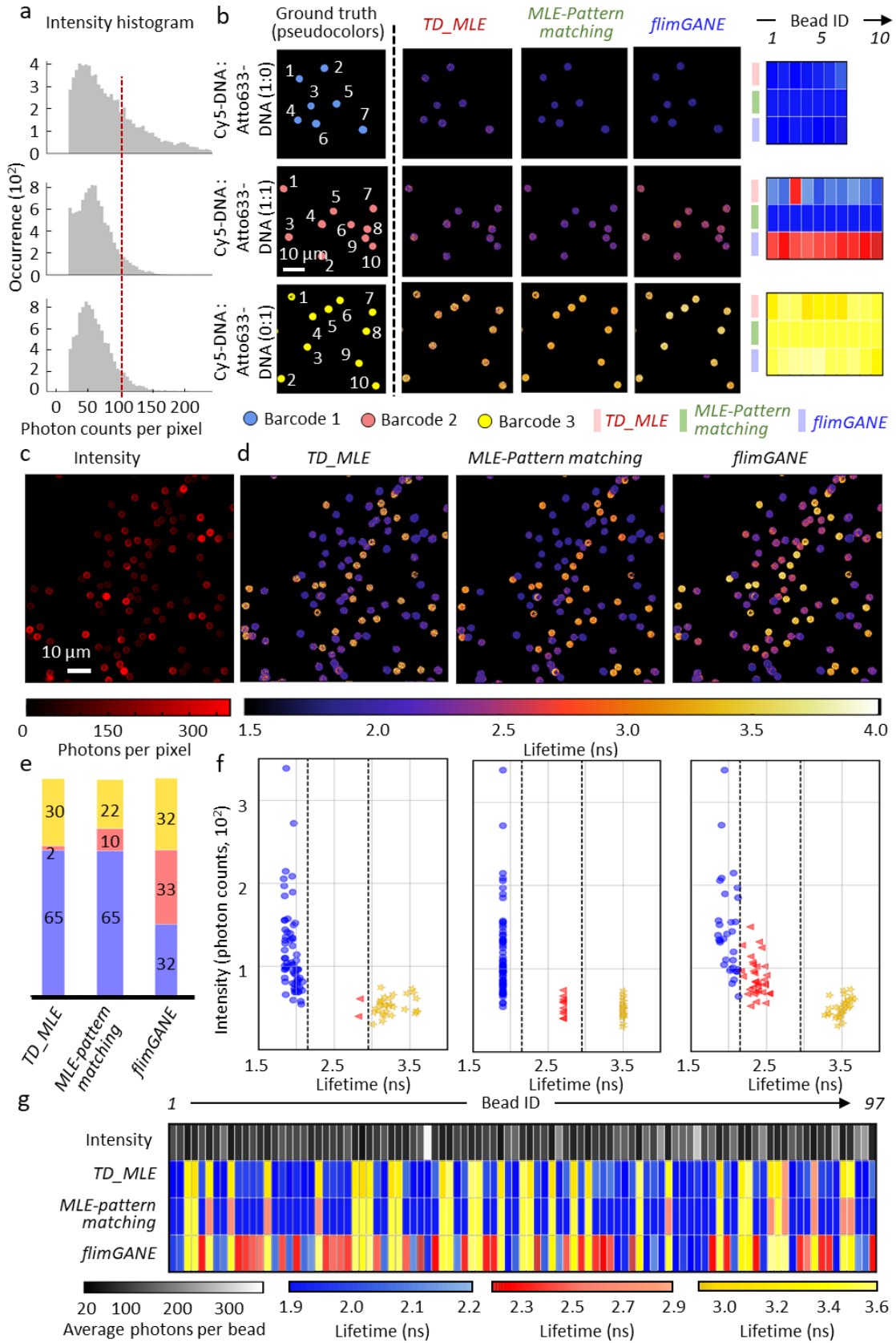
Supplementary Fig. 9 A real-world application of the CoME. Experimental fluorescence decay histogram aligned well with IRF based on CoME. Given the fluorescent dye with known lifetime (theoretical fluorescence lifetime of Atto633 = 3.3 ns), the desired difference of mass centers (unit: bins) of Atto633 was calculated. CoME analysis was then performed on the experimental IRF and fluorescence decay histogram to obtain the raw difference of CoM. Based on the calibration line, comparing the raw difference with the desired difference of CoM, the temporal shift can be determined with the calibration line, and the experimental fluorescence decay histogram can be reversely shifted back to the desired location. Then, shift was applied to all data collected in the same day.



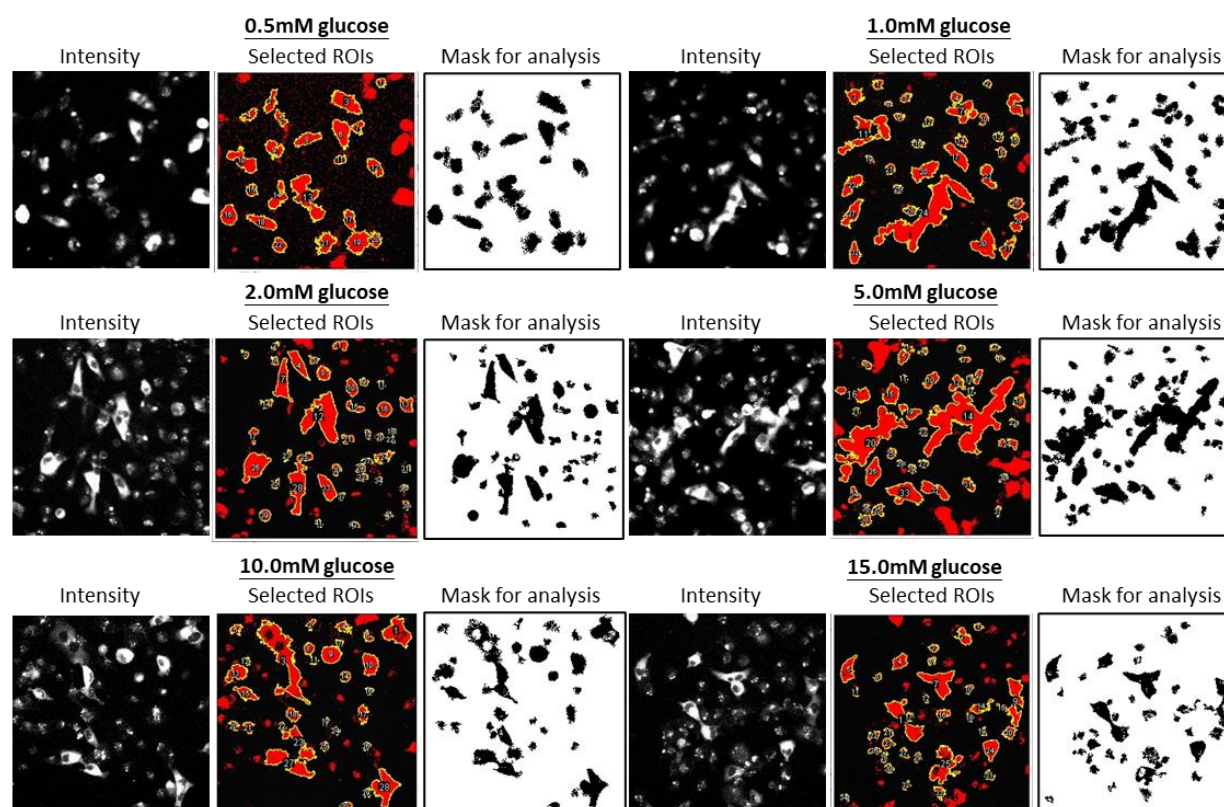
Supplementary Fig. 10 *flimGANE* can accurately discriminate fluorescence lifetime barcode beads individually at low-photon-count condition. **a-b** To better quantify the statistical result of FLIM barcode imaging **a**, we selected the appropriate ROIs **b**. Based on the intensity contrast image, we utilized ImageJ to obtain the ROIs for barcode bead. Then we manually adjusted some miss-selections and labelled each bead with a unique ID. The corresponding mean lifetime, size, and intensity information for each bead can then be extracted and recorded for further analysis. Here we can classify all beads into three different types of barcodes.



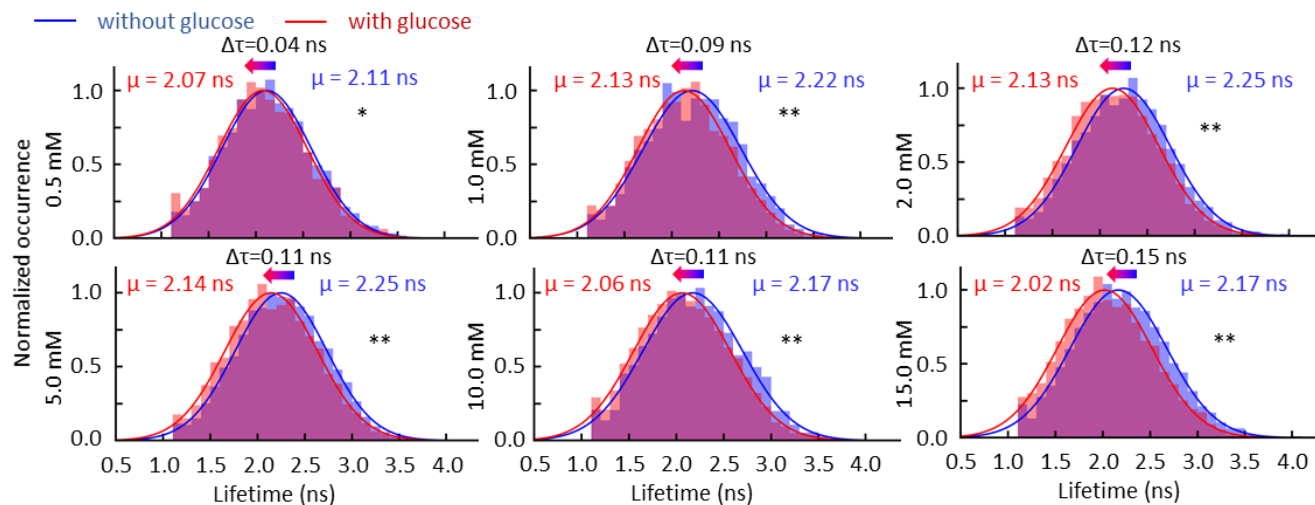
Supplementary Fig. 11 Two dimensional (intensity versus lifetime) scattered plots from different methods illustrated the success of *flimGANE* for identifying three populations of barcodes with nearly the same size. Each data point represents the statistical information of each bead analyzed by *TD_LSE*, *TD_MLE*, *DFD_LSE* and *flimGANE*, respectively. Mean intensity is derived by averaging the photon counts per pixel, and the mean lifetime is obtained from the Gaussian model fitting of apparent lifetime histogram generated by different analytical methods. Based on lifetime, beads are categorized into three groups: *barcode_1* (blue) for lifetime less than 2.15 ns, *barcode_2* (red) for lifetime lying within 2.15 and 2.95 ns, and *barcode_3* (yellow) for lifetime longer than 2.95 ns.



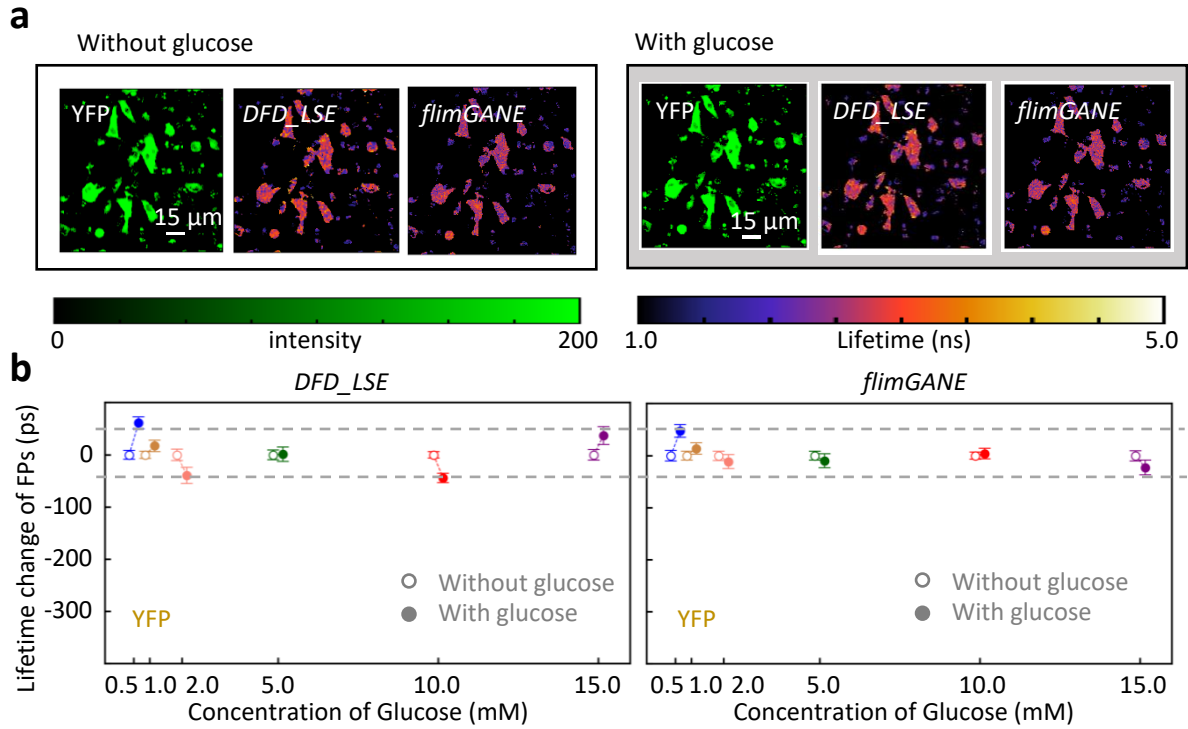
Supplementary Fig. 12 Performance comparison between *TD_MLE*, *MLE-pattern matching*, and *flimGANE* on fluorescence barcode classification. **a** Intensity histograms of three pure species imaging acquired individually. Red dashed line indicated 100 photon counts per pixel. **b** Ground-truth FLIM image and the FLIM images of each barcode were reconstructed by *TD_MLE*, *MLE-pattern matching*, and *flimGANE* (scale bar means 10 μm). The classification results were demonstrated with the same order as **g**. **c** A contrast image shows 3- μm fluorescent beads with various brightness, where **d** FLIM images obtained from *TD_MLE*, *MLE-pattern matching*, and *flimGANE* reveal three lifetime species. **e** Barcode classification results indicate that *MLE-pattern matching* is improved but still less accurate than *flimGANE*. **f** Scattered plots of intensity versus lifetime from different methods. **g** Classification results of all 97 beads for different methods.



Supplementary Fig. 13 CFP-g-YFP-transfected cells were selected from fluorescence image by ImageJ (ROI Manager). Since the signal-to-noise ratio of FLIM images for live cell FRET experiment was low, we had to performed ROI selection to quantify the fluorescence lifetime of CFP and YFP for the cells only by ImageJ, avoiding the misinterpretation introduced by the artifact. For each experimental condition (0.5, 1.0, 2.0, 5.0, 10.0, 15.0 mM glucose), we selected ROIs independently and generated the individual mask for further lifetime analysis.

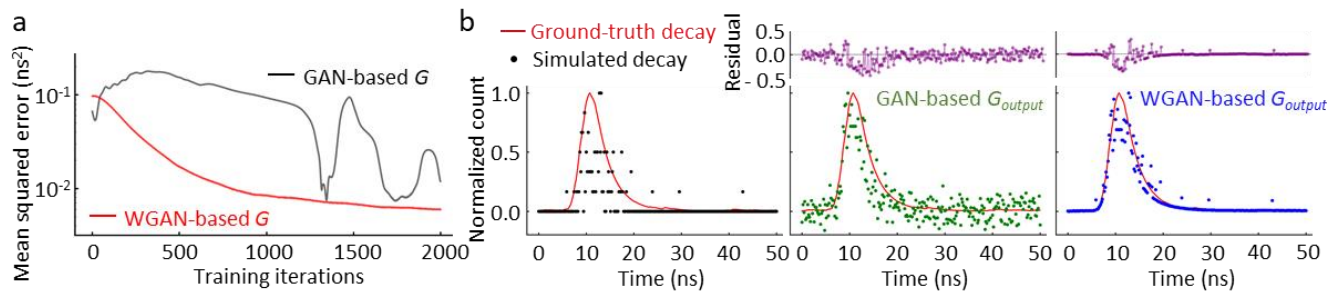


Supplementary Fig. 14 Normalized CFP lifetime distributions of CFP-g-TFP MDA-MB-231 cells and Gaussian fitting of each lifetime value for *flimGANE*. The CFP-g-YFP sensor-expressed MDA-MB-231 cell FLIM images were reconstructed by *flimGANE*. Based on the selected ROIs method from **Supplementary Fig. 13**, we extracted the fluorescence lifetime values from those ROIs and plotted these values in CFP distribution (blue: without glucose; red: with glucose). The mean value of lifetime for each histogram was extracted by Gaussian fitting. The fitting results were summarized in **Supplementary Table 12**. Significant differences are indicated as: ** (p -value < 0.001); * (p -value < 0.05) using two-sided KS (Kolmogorov-Smirnov) test.

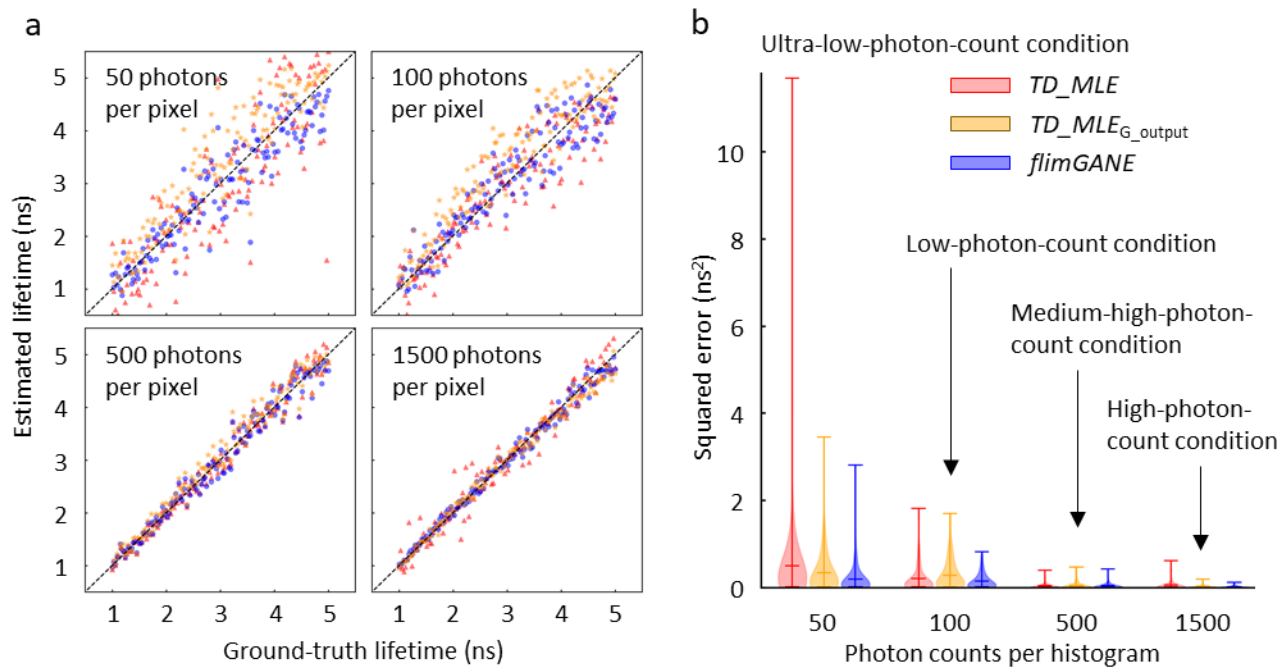


Supplementary Fig. 15 The quantification of CFP-g-YFP-transfected MDA-MB-231 cell FLIM images for YFP channel.

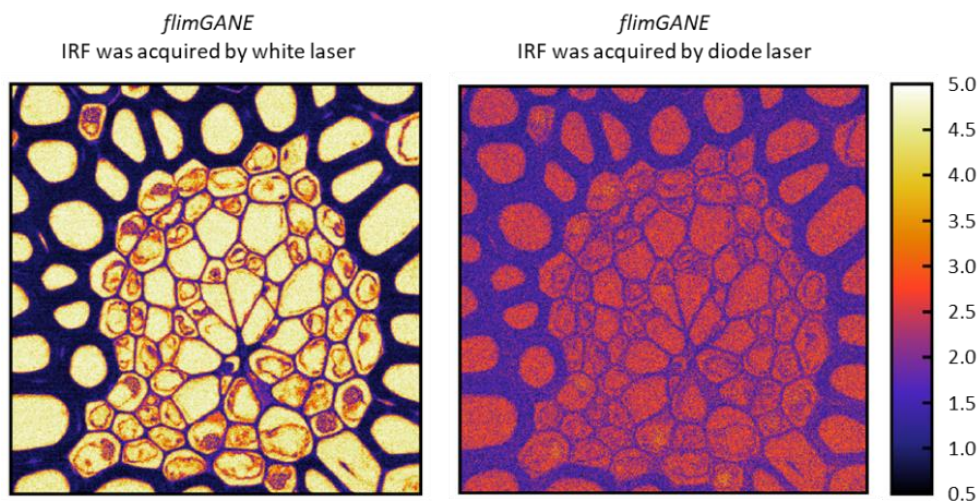
a CFP-g-YFP-transfected MDA-MB-231 cells FLIM images (w/o and w/ 2 mM glucose) for YFP channel were reconstructed by *flimGANE* and *DFD_LSE*. The images of live MDA-MB-231 cells incubated in 2 mM glucose was taken with the same field of view (FOV) as the images of cells incubated in culture medium without glucose. **b** The mean lifetime difference between the group without and with glucose was plotted versus the six concentrations of glucose (0.5, 1.0, 2.0, 5.0, 10.0, 15.0 mM). The variation of mean lifetime difference obtained by *flimGANE* before and after adding glucose was smaller ($< \pm 0.05$ ns) than that obtained by *DFD_LSE*. The error bars represented the coefficient of variation ($n = 2290\sim 6824$ pixels).



Supplementary Fig. 16 Comparison of WGAN and GAN on learning to reconstruct the ground-truth mimicking fluorescence decay histograms. **a** The mean squared error between G_{output} and the ground-truth fluorescence decay histograms over training iterations demonstrates that GAN training is unstable. In addition, the performance of GAN is worse than that of WGAN. **b** A representative example of simulated decay, GAN-based G_{output} , and WGAN-based G_{output} . Residual between G_{output} and ground-truth decay showing WGAN outperforms GAN on the reconstruction of the ground-truth fluorescence decay histograms.

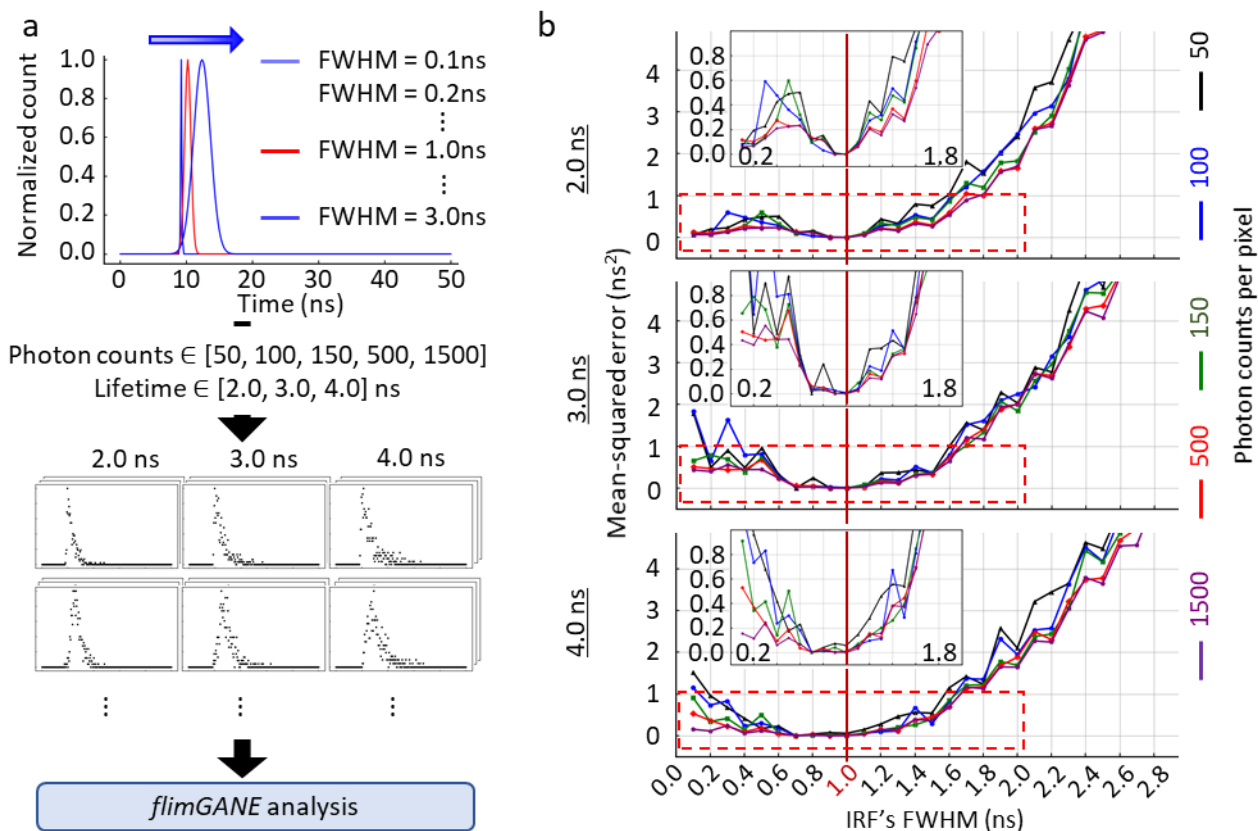


Supplementary Fig. 17 Comparison of TD_MLE tested on simulated decay histogram or on G_{output} ($TD_MLE_{G_output}$) and $flimGANE$. **a** 2D scatter plots of estimated and ground-truth lifetime. **b** Squared error with the ground truth comparison of different methods.

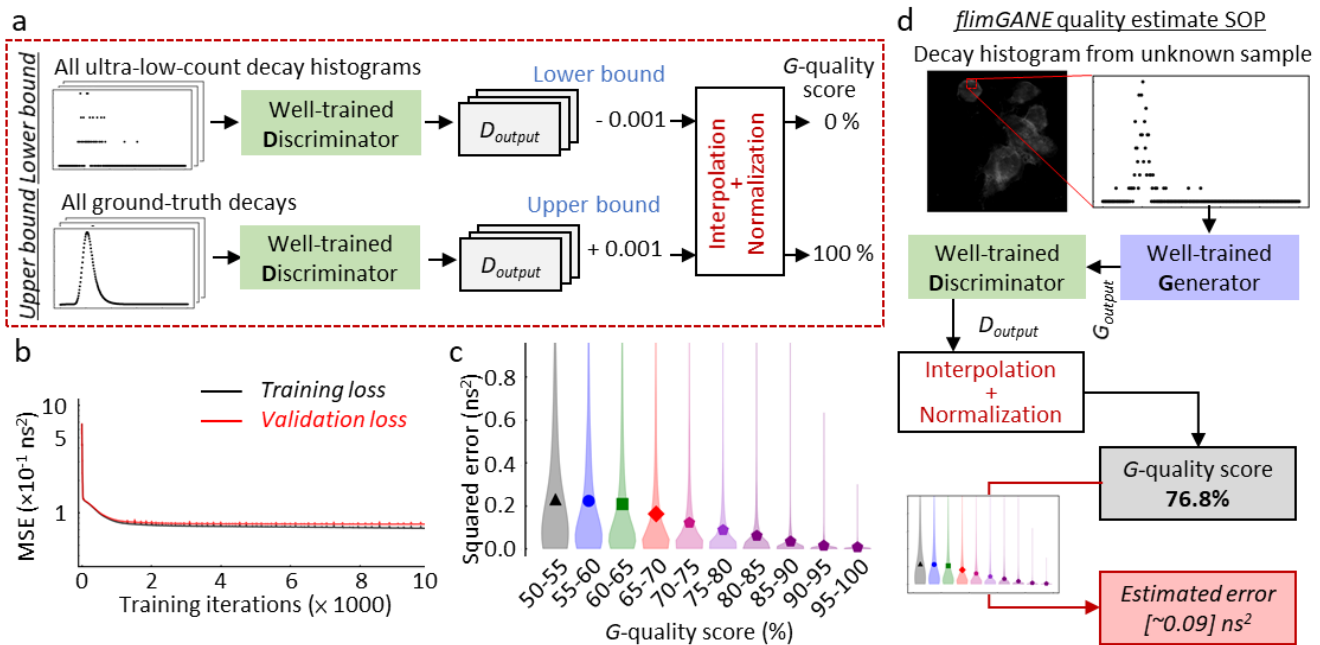


Supplementary Fig. 18 Incorrect *Convallaria* FLIM image was reconstructed by *flimGANE* with incorrect IRF input.

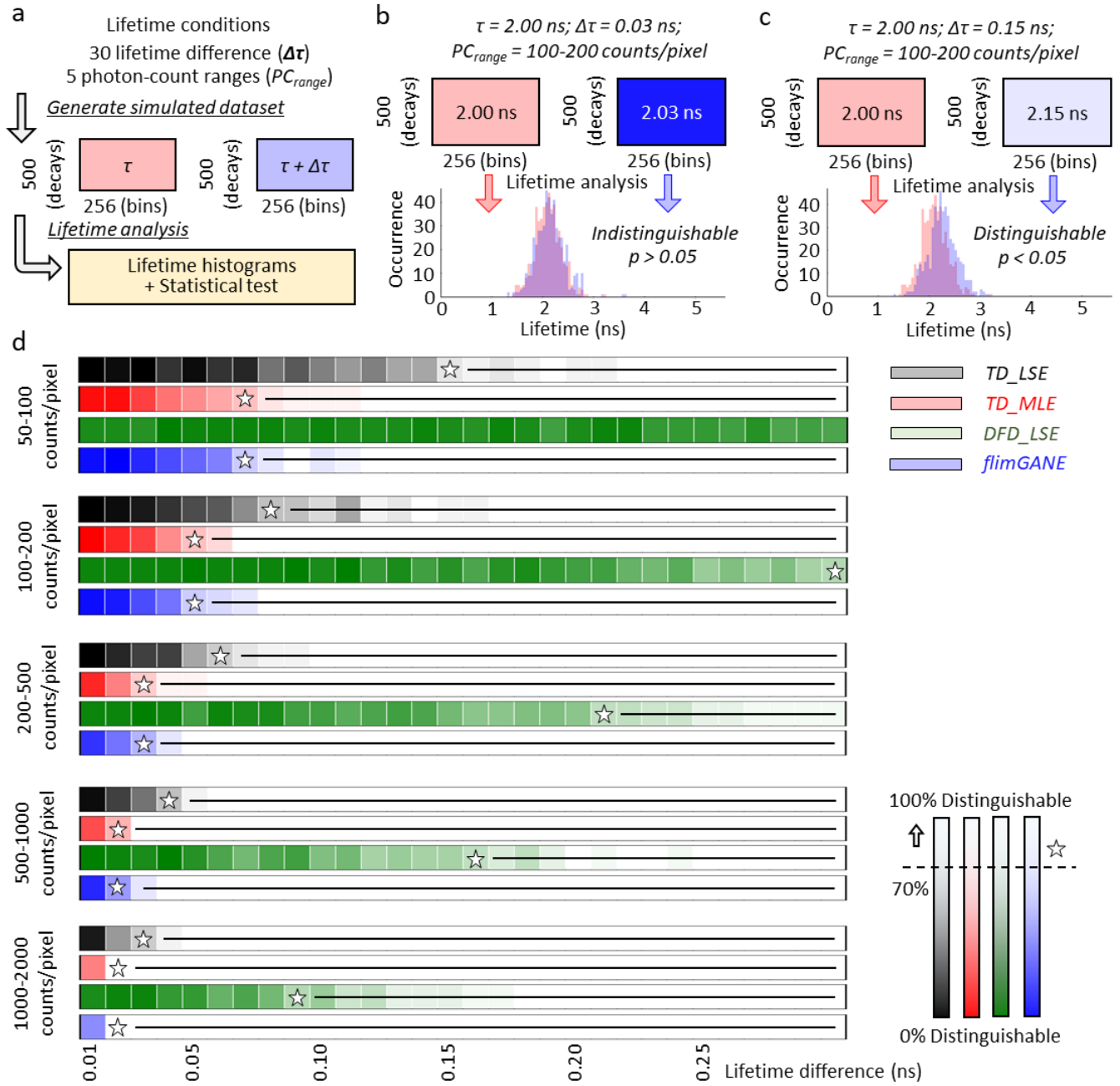
The IRF was determined by the laser, the detector and the FLIM system electronics used in experiments. To avoid complications, traditional software tools might use a synthesized IRF to perform the analysis. However, we want to emphasize that IRF should be considered as an important factor for fluorescence lifetime analysis. The IRF used for the left figure was acquired by white laser while the right one was acquired by diode laser. Although the shape visualized on the two images were similar, the estimated lifetime value for each pixel was different. It was obvious that the FLIM image reconstructed by incorrect IRF led to incorrect judgement for lifetime-based technique.



Supplementary Fig. 19 Influence of fluctuating IRF on *flimGANE* performance. **a** Workflow of simulation process to evaluate the performance over the variability of IRF. We generated several synthetic Gaussian IRFs with varying FWHMs (0.1, 0.2, ..., 3.0 ns). IRF with 1.0 ns FWHM (red) was utilized as a reference to generate the training dataset for *flimGANE* development. Each synthetic IRF was used to generate the testing dataset with multiple conditions, including different photon budgets (50, 100, 150, 500, 1500 photon counts per pixel) and lifetime value of interest (2.0, 3.0, 4.0 ns). This testing dataset was then fed into *flimGANE* for evaluation. **b** Mean-squared error between lifetime estimates and the ground truth (left) and the zoomed-in ROI (red dashed box, right) over disparate IRF's FWHM. Dark red vertical line represents the condition with the same IRF utilized for model training.



Supplementary Fig. 20 Quality estimate (G-quality score) in *flimGANE* based on known samples. **a** The lower and upper bounds of discriminator outputs (D_{output}) were determined by the minimum of D_{output} from all the ultra-low-count decay histograms (e.g., 50 photon counts per pixel) and maximum of D_{output} from all the ground-truth decays. We then mapped the D_{output} to G-quality score with the range of 0 and 100% via interpolation and normalization. **b** A typical plot of the E loss function of *flimGANE* combination training ($G+E$) process. Both the training loss and validation loss quickly converge to an equilibrium stage. **c** Dependency of the squared error of the lifetime estimate and the G-quality score showed that the higher G-quality corresponds to smaller lifetime estimation error. **d** Schematic of *flimGANE* quality estimate SOP given the input decay histogram from live HeLa cell sample.



Supplementary Fig. 21 Comparison on discriminability for TD_LSE , TD_MLE , DFD_LSE , and $flimGANE$. **a**

Discriminability test evaluated the statistical significance of differences between any two lifetime distributions. In this simulation, the lifetime difference ($\Delta\tau$) was ranging from 0.01 to 0.30 ns and the 5 photon-count conditions were 50-100, 100-200, and 200-500 photons per pixel. **b** Decays with $\Delta\tau = 0.03$ ns were indistinguishable as their p-value was 0.33 (> 0.05). **c** Decays with $\Delta\tau = 0.15$ ns were distinguishable as their p-value was 1.95×10^{-16} ($\ll 0.05$). **d** Heatmaps showing the discriminability ($\star$) of TD_LSE , TD_MLE , DFD_LSE , and $flimGANE$ under the 5 photon-count conditions. Using the p -value < 0.05 in two-sided KS test as a criterion, if the p -value was less than 0.05 for 70% of the repeated tests, the two distributions were called distinguishable.

Supplementary Table 1. Training database for different applications

Parameters	Fig. 1 (two dyes mixing in solution)	Fig. 2 (Barcode beads)	Fig. 3 (Convallari a)	Fig. 3 (HeLa cell) SI Fig. 15 (YFP)	Fig. 4 (CFP)	Fig. 5 (Auto- fluorescenc e – FAD)	Fig. 5 (Auto- fluorescenc e – NADH)
α_1	0.0 ~ 1.0	0.0 ~ 1.0	0.98 ~ 1.00	0.98 ~ 1.00	0.0 ~ 1.0	0.0 ~ 1.0	0.0 ~ 1.0
τ_1 (ns)	0.5 ~ 0.7	1.8 ~ 2.0	0.05 ~ 10.0	0.5 ~ 5.0	1.1	0.3	0.4
τ_2 (ns)	3.2 ~ 3.4	3.4 ~ 3.6	N/A	N/A	3.6	3.0	2.4 ~ 4.0
IRF with excitation laser	diode laser	diode laser	white laser	diode laser	diode laser	diode laser	diode laser
rate	50, 100, 500, 1,500, 5,000 photons per pixel						
No. of degraded decays	990 k	990 k	300 k	69 k	55 k	55 k	467.5 k
No. of ground truths	99	99	600	138	11	11	187
Training time for G and D (hrs)	18	18	6.1	1.5	1.4	1.4	6.2
Training time for E (hrs)	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Training time for combing G and E (hrs)	0.8	0.8	0.7	0.6	0.1	0.1	0.7
Total model training time (hrs)	18.9	18.9	6.9	2.2	1.6	1.6	7.0

Supplementary Table 2. Summary of different photon-count conditions

Ultra-low-photon- count	Low-photon-count	Medium-photon- count	Medium-high- photon-count	High-photon-count
50-100	100-200	200-500	500-1000	>1000

Supplementary Table 3. Mean squared error (MSE) comparison (mean $\pm$ s.e.m.; n = 1,340 pixels) of different analysis methods

Condition (photons per pixel)	<i>TD_LSE</i> (ns ²)	<i>TD_MLE</i> (ns ²)	<i>DFD_LSE</i> (ns ²)	<i>flimGANE</i> (ns ²)
50	0.91 $\pm$ 0.03	0.36 $\pm$ 0.01	1.65 $\pm$ 0.07	0.21 $\pm$ 0.01
60	0.66 $\pm$ 0.03	0.37 $\pm$ 0.02	1.20 $\pm$ 0.06	0.21 $\pm$ 0.01
80	0.63 $\pm$ 0.03	0.38 $\pm$ 0.02	1.04 $\pm$ 0.05	0.16 $\pm$ 0.01
100	0.54 $\pm$ 0.02	0.29 $\pm$ 0.02	0.59 $\pm$ 0.02	0.13 $\pm$ 0.01
150	0.34 $\pm$ 0.02	0.18 $\pm$ 0.01	0.52 $\pm$ 0.02	0.09 $\pm$ 0.00
500	0.12 $\pm$ 0.01	0.06 $\pm$ 0.00	0.20 $\pm$ 0.01	0.03 $\pm$ 0.00

1000	0.06 ± 0.00	0.03 ± 0.00	0.16 ± 0.01	0.02 ± 0.00
1500	0.04 ± 0.00	0.02 ± 0.00	0.09 ± 0.00	0.01 ± 0.00

Supplementary Table 4. Peak signal-to-noise ratio (PSNR) comparison (mean ± s.e.m.; n = 10 frames) of different analysis methods *in silico*

Condition (photons per pixel)	<i>TD_LSE</i>	<i>TD_MLE</i>	<i>DFD_LSE</i>	<i>flimGANE</i>
50	14.39 ± 0.13	18.49 ± 0.21	12.61 ± 0.80	20.79 ± 0.19
60	15.82 ± 0.12	18.38 ± 0.19	13.93 ± 0.71	20.80 ± 0.15
80	15.97 ± 0.09	18.20 ± 0.16	14.14 ± 0.51	22.06 ± 0.22
100	16.78 ± 0.29	19.42 ± 0.24	16.39 ± 0.29	22.88 ± 0.18
150	18.74 ± 0.20	21.49 ± 0.17	17.34 ± 0.62	24.34 ± 0.09
500	23.37 ± 0.25	26.55 ± 0.14	21.22 ± 0.50	29.23 ± 0.18
1000	26.39 ± 0.18	29.57 ± 0.23	22.52 ± 0.76	31.73 ± 0.15
1500	27.97 ± 0.11	30.77 ± 0.18	24.59 ± 0.54	33.14 ± 0.18

Supplementary Table 5. Structural similarity index (SSIM) comparison (mean ± s.e.m.; n = 10 frames) of different analysis methods *in silico*

Condition (photons per pixel)	<i>TD_LSE</i>	<i>TD_MLE</i>	<i>DFD_LSE</i>	<i>flimGANE</i>
50	0.94 ± 0.00	0.97 ± 0.00	0.83 ± 0.04	0.98 ± 0.00
60	0.95 ± 0.00	0.97 ± 0.00	0.88 ± 0.04	0.98 ± 0.00
80	0.96 ± 0.00	0.98 ± 0.00	0.91 ± 0.02	0.99 ± 0.00
100	0.96 ± 0.00	0.98 ± 0.00	0.95 ± 0.00	0.99 ± 0.00
150	0.98 ± 0.00	0.99 ± 0.00	0.96 ± 0.01	0.99 ± 0.00
500	0.99 ± 0.00	1.00 ± 0.00	0.98 ± 0.00	1.00 ± 0.00
1000	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.00	1.00 ± 0.00
1500	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.00	1.00 ± 0.00

Supplementary Table 6. Goodness-of-fit reduced χ^2 comparison (mean ± s.e.m.; n = 1,340 pixels) of different analysis methods *in silico* (mean ± standard deviation)

Condition (photons per pixel)	<i>TD_LSE</i>	<i>TD_MLE</i>	<i>DFD_LSE</i>	<i>flimGANE</i>
50	0.58 ± 0.00	0.62 ± 0.00	0.83 ± 0.01	N/A
60	0.60 ± 0.00	0.64 ± 0.00	0.83 ± 0.01	N/A
80	0.67 ± 0.00	0.73 ± 0.01	0.82 ± 0.02	N/A
100	0.74 ± 0.00	1.01 ± 0.03	0.74 ± 0.01	N/A
150	0.88 ± 0.01	0.95 ± 0.01	0.70 ± 0.01	N/A

500	0.98 ± 0.01	1.04 ± 0.01	0.56 ± 0.01	N/A
1000	0.96 ± 0.00	1.00 ± 0.00	0.48 ± 0.02	N/A
1500	0.92 ± 0.00	0.95 ± 0.00	0.39 ± 0.01	N/A

Supplementary Table 7. Execution time comparison of different analysis methods *in silico*

<i>TD_LSE</i> (ms / pixel)	<i>TD_MLE</i> (ms / pixel)	<i>DFD_LSE</i> (ms / pixel)	<i>flimGANE</i> (ms / pixel)
82.40	906.37	3.94	0.32

Supplementary Table 8. Summary of different analysis methods for the mixture of two fluorescent dyes (* mean $\pm$ standard deviation from Gaussian distributing fitting)

Ratio (Cy5:Atto 633)	Photons per pixel	Theoretical apparent lifetime, τ_a (ns)	<i>TD_LSE*</i> (ns)	<i>TD_MLE*</i> (ns)	<i>DFD_LSE*</i> (ns)	<i>flimGANE*</i> (ns)	<i>flimGANE</i> (Pre- exponential factor, α_1)*
10:0	178	0.60	0.59 ± 0.05	0.71 ± 0.03	0.62 ± 0.02	0.68 ± 0.15	0.98 ± 0.05
9:1	135	0.87	0.81 ± 0.14	1.11 ± 0.17	0.70 ± 0.06	0.88 ± 0.10	0.90 ± 0.05
8:2	127	1.14	1.05 ± 0.19	1.28 ± 0.20	0.80 ± 0.09	1.13 ± 0.11	0.81 ± 0.05
7:3	121	1.41	1.02 ± 0.19	1.52 ± 0.23	0.95 ± 0.13	1.42 ± 0.14	0.70 ± 0.05
6:4	127	1.68	1.10 ± 0.22	1.82 ± 0.27	1.09 ± 0.21	1.70 ± 0.17	0.59 ± 0.05
5:5	120	1.95	1.53 ± 0.26	1.92 ± 0.26	1.21 ± 0.28	1.99 ± 0.25	0.48 ± 0.10
4:6	129	2.22	1.68 ± 0.27	2.32 ± 0.33	1.29 ± 0.29	2.26 ± 0.25	0.39 ± 0.08
3:7	134	2.49	1.74 ± 0.28	2.49 ± 0.34	1.99 ± 1.20	2.64 ± 0.37	0.27 ± 0.10
2:8	137	2.76	1.80 ± 0.33	2.85 ± 0.37	2.14 ± 1.27	2.82 ± 0.20	0.18 ± 0.07
0:10	190	3.30	2.67 ± 0.60	3.32 ± 0.06	3.04 ± 1.05	3.31 ± 0.30	0.00 ± 0.13

Supplementary Table 9. Summary of each detected bead

ID	<i>TD_LSE</i> τ_a (ns)	<i>TD_LSE</i> (Barcode)	<i>TD_MLE</i> τ_a (ns)	<i>TD_MLE</i> (Barcode)	<i>DFD_LSE</i> τ_a (ns)	<i>DFD_LSE</i> (Barcode)	<i>flimGANE</i> τ_a (ns)	<i>flimGANE</i> (Barcode)
1	1.83	1	1.85	1	1.93	1	1.96	1
2	1.83	1	1.86	1	1.97	1	2.02	1
3	2.84	2	3.26	3	3.52	3	3.47	3
4	2.85	2	3.11	3	3.52	3	3.52	3
5	1.87	1	2.00	1	2.12	1	2.31	2
6	2.94	2	3.23	3	3.52	3	3.54	3
7	1.83	1	1.85	1	1.96	1	2.05	1
8	1.84	1	1.84	1	1.97	1	2.09	1
9	2.85	2	3.16	3	3.52	3	3.48	3

10	1.85	1	1.98	1	2.10	1	2.33	2
11	1.87	1	1.97	1	2.07	1	2.38	2
12	1.90	1	1.97	1	2.13	1	2.47	2
13	1.88	1	1.99	1	2.19	2	2.53	2
14	2.89	2	3.13	3	3.52	3	3.54	3
15	1.83	1	1.84	1	1.94	1	2.07	1
16	1.88	1	1.95	1	2.10	1	2.43	2
17	1.84	1	1.84	1	1.94	1	2.05	1
18	1.87	1	1.93	1	2.06	1	2.32	2
19	1.83	1	1.84	1	1.95	1	2.06	1
20	1.84	1	1.86	1	1.95	1	2.13	1
21	2.90	2	3.11	3	3.52	3	3.61	3
22	1.87	1	1.98	1	2.10	1	2.38	2
23	1.85	1	1.97	1	2.11	1	2.50	2
24	1.88	1	1.95	1	2.11	1	2.48	2
25	1.87	1	1.92	1	2.06	1	2.40	2
26	2.87	2	3.18	3	3.52	3	3.58	3
27	2.68	2	3.01	3	3.10	3	3.45	3
28	2.93	2	3.28	3	3.52	3	3.63	3
29	1.87	1	2.01	1	2.11	1	2.48	2
30	1.84	1	1.85	1	1.95	1	2.06	1
31	2.78	2	3.12	3	3.52	3	3.54	3
32	2.84	2	3.23	3	3.52	3	3.58	3
33	1.89	1	2.05	1	2.15	2	2.49	2
34	1.84	1	1.87	1	1.96	1	2.12	1
35	1.84	1	1.85	1	1.99	1	2.17	2
36	1.85	1	1.86	1	2.03	1	1.94	1
37	1.86	1	2.04	1	2.13	1	2.28	2
38	2.84	2	3.09	3	3.52	3	3.50	3
39	2.99	3	3.59	3	3.52	3	3.49	3
40	1.83	1	1.85	1	1.95	1	2.10	1
41	1.89	1	2.02	1	2.14	1	2.35	2
42	2.93	2	3.59	3	3.52	3	3.58	3
43	3.08	3	3.53	3	3.52	3	3.52	3
44	1.91	1	2.00	1	2.11	1	2.33	2

45	1.88	1	1.96	1	2.10	1	2.34	2
46	2.75	2	3.13	3	3.52	3	3.51	3
47	1.87	1	1.95	1	2.08	1	1.88	1
48	1.84	1	1.85	1	1.96	1	2.04	1
49	3.57	3	3.55	3	3.52	3	3.54	3
50	2.88	2	3.17	3	3.52	3	3.49	3
51	1.89	1	1.99	1	2.11	1	2.19	2
52	1.84	1	1.87	1	1.94	1	2.03	1
53	2.70	2	3.03	3	3.03	3	3.31	3
54	1.83	1	1.84	1	1.95	1	2.12	1
55	1.90	1	2.05	1	2.14	1	2.47	2
56	2.75	2	3.11	3	3.52	3	3.42	3
57	1.85	1	2.01	1	2.13	1	2.35	2
58	2.78	2	3.15	3	3.52	3	3.43	3
59	1.93	1	2.08	1	2.13	1	2.38	2
60	1.87	1	2.02	1	2.08	1	2.25	2
61	1.91	1	2.08	1	2.13	1	2.44	2
62	1.88	1	1.95	1	2.09	1	1.87	1
63	1.84	1	1.86	1	1.95	1	1.93	1
64	1.89	1	2.02	1	2.10	1	2.17	2
65	3.01	3	3.57	3	3.52	3	3.27	3
66	1.88	1	1.98	1	2.11	1	1.88	1
67	1.87	1	2.02	1	2.12	1	2.11	1
68	1.87	1	1.94	1	2.07	1	1.90	1
69	3.46	3	3.43	3	3.52	3	3.49	3
70	1.84	1	1.88	1	1.95	1	1.87	1
71	1.85	1	1.88	1	1.95	1	1.98	1
72	1.86	1	1.92	1	2.03	1	1.86	1
73	1.89	1	1.96	1	2.13	1	1.90	1
74	1.85	1	1.87	1	1.93	1	1.94	1
75	1.88	1	1.97	1	2.12	1	2.28	2
76	2.79	2	3.04	3	3.22	3	3.51	3
77	1.91	1	2.01	1	2.12	1	2.32	2
78	1.84	1	1.85	1	1.97	1	2.03	1
79	2.96	3	3.28	3	3.52	3	3.33	3

80	3.01	3	3.60	3	3.52	3	3.33	3
81	1.89	1	2.00	1	2.16	2	2.29	2
82	1.89	1	1.95	1	2.12	1	1.95	1
83	2.87	2	3.17	3	3.52	3	3.66	3
84	2.92	2	3.24	3	3.52	3	3.61	3
85	2.59	2	2.84	2	3.08	3	3.63	3
86	1.83	1	1.85	1	1.97	1	1.88	1
87	1.87	1	1.95	1	2.07	1	2.37	2
88	1.89	1	2.02	1	2.12	1	2.54	2
89	2.57	2	2.82	2	3.09	3	3.42	3
90	1.87	1	1.98	1	2.04	1	2.35	2
91	1.86	1	1.96	1	2.01	1	2.36	2
92	1.83	1	1.83	1	1.96	1	1.88	1
93	3.00	3	3.19	3	3.52	3	3.51	3
94	3.40	3	3.46	3	3.53	3	3.34	3
95	1.83	1	1.84	1	1.95	1	2.12	1
96	1.83	1	1.86	1	1.98	1	2.06	1
97	1.90	1	2.06	1	2.17	2	2.29	2

Supplementary Table 10. MSE, PSNR, SSIM, and VIF comparison of different analysis methods for *Convallaria* FLIM images (Standard: medium-count *TD_MLE* FLIM)

Condition (photons per pixel)	Method	MSE	PSNR	SSIM	VIF
50 – 150 photon counts per pixel	<i>TD_LSE</i>	2.01	10.94	0.71	0.17
	<i>TD_MLE</i>	0.52	16.84	0.89	0.22
	<i>DFD_LSE</i>	4.10	7.85	0.51	0.09
	<i>flimGANE</i>	0.61	15.71	0.88	0.22

Supplementary Table 11. MSE, PSNR, SSIM and VIF comparison of different analysis methods for live HeLa cell FLIM images (Standard: medium-high-count FLIM)

Condition (average photons per pixel)	Method	MSE	PSNR	SSIM	VIF
100 – 500 (Red channel)	<i>TD_LSE</i>	0.28	19.44	0.67	0.06
	<i>TD_MLE</i>	0.10	23.90	0.75	0.14
	<i>DFD_LSE</i>	0.24	20.26	0.68	0.11
	<i>flimGANE</i>	0.13	22.88	0.87	0.22

100 – 400 (Blue channel)	<i>TD_LSE</i>	0.37	18.34	0.93	0.25
	<i>TD_MLE</i>	0.12	23.10	0.95	0.45
	<i>DFD_LSE</i>	0.79	15.01	0.93	0.31
	<i>flimGANE</i>	0.10	24.01	0.98	0.47

Supplementary Table 12. Apparent lifetime of CFP calculated with Gaussian distribution fitting (* mean $\pm$ standard deviation errors on the parameter estimate)

FP type	[Glucose] (mM)	With or without glucose	<i>TD_LSE</i> *	<i>TD_MLE</i> *	<i>DFD_LSE</i> *	<i>flimGANE</i> *
CFP	0.5	Before	1.62 $\pm$ 0.02	2.08 $\pm$ 0.01	N/A	2.11 $\pm$ 0.01
		After	1.58 $\pm$ 0.02	2.08 $\pm$ 0.01	N/A	2.07 $\pm$ 0.01
	1.0	Before	1.63 $\pm$ 0.03	2.12 $\pm$ 0.01	N/A	2.22 $\pm$ 0.01
		After	1.59 $\pm$ 0.02	2.06 $\pm$ 0.01	N/A	2.13 $\pm$ 0.01
	2.0	Before	1.72 $\pm$ 0.02	2.19 $\pm$ 0.01	N/A	2.25 $\pm$ 0.01
		After	1.54 $\pm$ 0.02	2.05 $\pm$ 0.01	N/A	2.13 $\pm$ 0.01
	5.0	Before	1.62 $\pm$ 0.02	2.18 $\pm$ 0.01	N/A	2.25 $\pm$ 0.00
		After	1.46 $\pm$ 0.02	2.02 $\pm$ 0.01	N/A	2.14 $\pm$ 0.01
	10.0	Before	1.64 $\pm$ 0.04	2.01 $\pm$ 0.01	N/A	2.17 $\pm$ 0.01
		After	1.55 $\pm$ 0.04	1.95 $\pm$ 0.01	N/A	2.06 $\pm$ 0.01
	15.0	Before	1.62 $\pm$ 0.03	2.07 $\pm$ 0.01	N/A	2.17 $\pm$ 0.01
		After	1.55 $\pm$ 0.02	2.01 $\pm$ 0.01	N/A	2.02 $\pm$ 0.01

Supplementary Table 13. MSE, PSNR and SSIM comparison of different analysis methods for FRET FLIM images (Reference: *TD_MLE* FLIM)

FP type	[Glucose] (mM)	With or without glucose	Method	MSE	PSNR	SSIM
CFP	0.5	Before	<i>TD_LSE</i>	0.69	15.59	0.98
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.95	14.20	0.98
			<i>flimGANE</i>	0.29	19.40	0.99
		After	<i>TD_LSE</i>	0.66	15.77	0.98
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.90	14.43	0.98
			<i>flimGANE</i>	0.26	19.76	0.99
	1.0	Before	<i>TD_LSE</i>	0.61	16.13	0.98
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.90	14.46	0.97

		After	<i>flimGANE</i>	0.27	19.68	0.99
			<i>TD_LSE</i>	0.56	16.48	0.98
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.84	14.75	0.97
			<i>flimGANE</i>	0.25	20.09	0.99
	2.0	Before	<i>TD_LSE</i>	0.57	16.41	0.97
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.89	14.49	0.96
			<i>flimGANE</i>	0.26	19.79	0.98
		After	<i>TD_LSE</i>	0.51	16.94	0.97
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.80	14.94	0.96
			<i>flimGANE</i>	0.18	21.36	0.99
	5.0	Before	<i>TD_LSE</i>	0.55	16.58	0.94
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.32	18.93	0.95
			<i>flimGANE</i>	0.21	20.78	0.97
		After	<i>TD_LSE</i>	0.48	17.19	0.95
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.75	15.21	0.93
			<i>flimGANE</i>	0.14	22.57	0.98
	10.0	Before	<i>TD_LSE</i>	0.45	17.43	0.98
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.85	14.67	0.98
			<i>flimGANE</i>	0.25	20.07	0.99
		After	<i>TD_LSE</i>	0.43	17.61	0.99
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.85	14.68	0.98
			<i>flimGANE</i>	0.22	20.51	0.99
	15.0	Before	<i>TD_LSE</i>	0.53	16.76	0.97
			<i>TD_MLE</i>	N/A		
			<i>DFD_LSE</i>	0.84	14.73	0.96
			<i>flimGANE</i>	0.25	20.01	0.99
		After	<i>TD_LSE</i>	0.49	17.06	0.98
			<i>TD_MLE</i>	N/A		

			<i>DFD_LSE</i>	0.81	14.87	0.97
			<i>flimGANE</i>	0.24	20.11	0.99

Supplementary Table 14. MSE, PSNR, and SSIM comparison of different methods for autofluorescence FLIM images
(Reference: *TD_MLE* FLIM)

Condition	Method	MSE	PSNR	SSIM
FAD	<i>TD_LSE</i>	0.22	20.48	0.88
	<i>DFD_LSE</i>	0.69	15.59	0.71
	<i>flimGANE</i>	0.24	20.08	0.86
NADH	<i>TD_LSE</i>	0.66	15.76	0.66
	<i>DFD_LSE</i>	1.39	12.55	0.57
	<i>flimGANE</i>	0.21	20.65	0.87

Supplementary References

- Goodfellow I, *et al.* Generative adversarial nets. In: *Advances in Neural Information Processing Systems*) (2014).
- Farnia F, Ozdaglar A. GANs may have no nash equilibria. *arXiv preprint at <https://arxiv.org/abs/2002.09124>*, (2020).
- Su J. GAN-QP: A novel GAN framework without gradient vanishing and lipschitz constraint. *arXiv preprint at <https://arxiv.org/abs/181.107296>*, (2018).
- Arjovsky M, Chintala S, Bottou L. Wasserstein gan. *arXiv preprint at <https://arxiv.org/abs/1701.07875>*, (2017).
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*) (2016).
- Chen W, Hays J. Sketchygan: Towards diverse and realistic sketch to image synthesis. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*) (2018).
- Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint at <https://arxiv.org/abs/1412.6980>*, (2014).
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600-612 (2004).
- Korhonen J, You J. Peak signal-to-noise ratio revisited: Is simple beautiful? In: *IEEE Fourth International Workshop on Quality of Multimedia Experience*) (2012).
- Sheikh HR, Bovik AC. A visual information fidelity approach to video quality assessment. In: *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*) (2005).
- Lakowicz JR, Laczko G, Cherek H, Gratton E, Limkeman M. Analysis of fluorescence decay kinetics from variable-frequency phase shift and modulation data. *Biophysical Journal* **46**, 463 (1984).
- Sun Y, Day RN, Periasamy A. Investigating protein-protein interactions in living cells using fluorescence lifetime

- imaging microscopy. *Nature Protocols* **6**, 1324 (2011).
13. Schawinski K, Zhang C, Zhang H, Fowler L, Santhanam GK. Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Monthly Notices of the Royal Astronomical Society: Letters* **467**, L110-L114 (2017).
 14. Rivenson Y, *et al.* Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature Biomedical Engineering* **3**, 466 (2019).
 15. Wang H, *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110 (2019).
 16. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. *arXiv preprint at <https://arxiv.org/abs/1705.10843>*, (2017).
 17. Yang Q, *et al.* Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging* **37**, 1348-1357 (2018).
 18. Ni K, Bresson X, Chan T, Esedoglu S. Local histogram based segmentation using the Wasserstein distance. *International Journal of Computer Vision* **84**, 97-111 (2009).
 19. Yu D, Yao K, Su H, Li G, Seide F. KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE (2013).