HHS Public Access

Author manuscript

Nat Photonics. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Nat Photonics. 2023 March ; 17(3): 250–258. doi:10.1038/s41566-022-01140-6.

Artificial confocal microscopy for deep label-free imaging

Xi Chen[1,2], **Mikhail E. Kandel**[1,3], **Shenghua He**[4], **Chenfei Hu**[1,5], **Young Jae Lee**[1,6], **Kathryn Sullivan**[7], **Gregory Tracy**[8], **Hee Jung Chung**[6,8], **Hyun Joon Kong**[1,7,9,10], **Mark Anastasio**[1,7], **Gabriel Popescu**[1,5,7,10]

[1]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[2]Currently with School of Applied and Engineering Physics, Cornell University, Ithaca, USA

[3]Currently with Groq, 400 Castro St., Suite 600, Mountain View, CA 94041, USA

[4]Department of Computer Science & Engineering, Washington University in St. Louis, St. Louis, Missouri, 63130, USA

[5]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[6]Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[7]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[8]Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

[9]Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[10]Carl Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Abstract

Widefield microscopy of optically thick specimens typically features reduced contrast due to “spatial crosstalk”, in which the signal at each point in the field of view is the result of a superposition from neighbouring points that are simultaneously illuminated. In 1955, Marvin Minsky proposed confocal microscopy as a solution to this problem. Today, laser scanning confocal fluorescence microscopy is broadly used due to its high depth resolution and sensitivity, but comes at the price of photobleaching, chemical, and photo-toxicity. Here, we present artificial

Corresponding author: Xi Chen, Cornell University, 148 Clark Hall, 142 Sciences Dr, Ithaca, NY 14853, xc289@cornell.edu.

Author Contributions Statement: X.C., M.E.K., and G.P. conceived the project. X.C. and M.E.K. designed the experiments. X.C. and M.E.K. built the system. X.C. performed imaging. S.H. trained the machine learning network. X.C. and M.E.K. analyzed the data. G.T & H.J.C. provided neurons. Y.J.L. cultured neurons and performed immunocytochemistry. K.M.S. & H.K. provided spheroids. X.C., C.H., and G.P. derived the theoretical model. X.C., M.E.K., S.H., C.H., and G.P. wrote the manuscript. M.A. supervised the AI work. G.P. supervised the project.

Code availability:

The codes that support the findings of this study are available from the corresponding author upon reasonable request.

Competing Interests Statement: G.P. had a financial interest in Phi Optics, Inc., a company developing quantitative-phase-imaging technology for materials and life science applications. The remaining authors declare that they have no conflict of interest.

confocal microscopy (ACM) to achieve confocal-level depth sectioning, sensitivity, and chemical specificity, on unlabeled specimens, nondestructively. We equipped a commercial laser scanning confocal instrument with a quantitative phase imaging module, which provides optical path-length maps of the specimen in the same field of view as the fluorescence channel. Using pairs of phase and fluorescence images, we trained a convolution neural network to translate the former into the latter. The training to infer a new tag is very practical as the input and ground truth data are intrinsically registered, and the data acquisition is automated. The ACM images present significantly stronger depth sectioning than the input (phase) images, enabling us to recover confocal-like tomographic volumes of microspheres, hippocampal neurons in culture, and 3D liver cancer spheroids. By training on nucleus-specific tags, ACM allows for segmenting individual nuclei within dense spheroids for both cell counting and volume measurements. In summary, ACM can provide quantitative, dynamic data, nondestructively from thick samples, while chemical specificity is recovered computationally.

Introduction

Three-dimensional (3D) cellular systems have been increasingly adopted over 2D cell monolayers to study disease mechanisms[1] and discover drug therapeutics[2], as they more accurately recapitulate the *in vivo* cellular function and development of extracellular matrices[3]. 3D cellular structures, including cellular clusters such as organoids and spheroids, have found use in a wide range of applications such as tissue engineering[4], high-throughput toxicology[5], and personalized medicine[6]. A particularly exciting direction of research is engineering multicellular living systems[7–9]. These fields of current scientific interest bring along the urgent need for new methods of investigation to inform about cellular viability and cell cluster proliferation. Ideally, such techniques would provide quantitative data with subcellular resolution at arbitrary depths in the cellular system and dynamic information rendered over broad time scales. Importantly, these assays would be completely nondestructive, i.e., would report on the cell cluster without interfering with its viability and function.

Due to the visible light wavelengths being comparable with the subcellular structures, optical methods of investigation are well suited for meeting these requirements. However, significant challenges exist for the existing optical microscopy techniques when applied to increasingly thick samples. Typical spheroids, ranging in size from hundreds of microns to millimeters, are significantly larger than the scattering mean free path associated with the light wave propagation, generate strong multiple scattering, and, thus, form optically turbid aggregates that are difficult to analyze at a cellular level[10,11]. As a result, high-throughput investigations often are limited to extracting coarse parameters, such as spheroid diameters, at low-magnification[12].

In 1955, in his pursuit to image 3D biological neuronal networks and mimic their behavior computationally, Minsky was faced with the challenge of suppressing multiple scattering, which was particularly severe for the widefield instruments available at the time[13]. In Minsky's own words, "One day it occurred to me that the way to avoid all that scattered light was to never allow any unnecessary light to enter in the first place. An ideal

microscope would examine each point of the specimen and measure the amount of light scattered or absorbed by that point."[13] This first implementation of the confocal scanning microscope was established in a transmission geometry, requiring sample translation. Of course, today's modern confocal instruments take advantage of bright laser sources, use beam scanning, and are most often used in a reflection geometry, paired with fluorescence contrast[14]. In time, many other advanced laser scanning techniques have been developed for fluorescence microscopy[15]. Nevertheless, fluorescence imaging is subject to several limitations. Absorption of the excitation light may cause the fluorophore to photobleach, which limits the time interval over which continuous imaging can be performed[16]. The excitation light is typically toxic to cells, a phenomenon referred to as phototoxicity, while the exogenous fluorophores themselves can induce chemical toxicity[17]. While the advancement of the green fluorescent protein technology significantly improves the viability of the specimen under investigation, concerns regarding phototoxicity, photobleaching, and functional integrity of the cells upon genetic engineering still remain[18]. Overcoming these limitations becomes extremely challenging when imaging thick objects over an extended period of time and, for that reason, confocal microscopy is often used on fixed specimens[19,20].

Multiphoton techniques, including harmonic generation and two-photon (intrinsic) fluorescence microscopy, have been established as valuable label-free for deep-tissue imaging with cellular resolution. Multiphoton microscopy (MPM) uses excitation light with a longer wavelength that penetrates deeper into tissues, while the nonlinear process requires a multiphoton interaction that renders three-dimensional localized excitation[21]. However, MPM requires expensive instrumentations such as femtosecond lasers that are less accessible to the broader community, and the higher-order nonlinear excitation is more susceptible to focus aberrations and phototoxicity. Light sheet fluorescence microscopy can acquire 3D tomography of biological specimens in seconds with high optical sectioning and axial resolution due to the sheet-like illumination, minimizing the background fluorescence and photobleaching[22]. However, tomographic reconstruction often requires sample rotation and sophisticated sample mounting. On the other hand, optical coherence tomography (OCT), an interferometric label-free method, was reported to detect and count aqueous cells in the anterior chamber of a rodent model of eye inflammation[23] and volumetrically quantify tumor spheroids[24]. Several phase-sensitive methods developed in a confocal modality have been recently developed, but their application to thick structures has been mostly unexplored[25–28].

Quantitative phase imaging (QPI)[29] has emerged recently as a potentially valuable label-free approach, which, due to its high resolution and sensitivity, has found a broad range of new applications[30]. While most of QPI's applications involve thin specimens (cell monolayers, thin tissue slices)[31–33], several efforts have been made for thick multiple-scattering samples. A multi-layer Born model using a first Born approximation at each of many layers has been proposed to provide phase reconstruction for thick samples[34]. Nevertheless, the maximum thickness presented in this model was around 30 μm, and the performance for 3D organoids with thickness usually on the order of 100 μm remains unclear. A multiscale reconstruction and stitching algorithm for optical diffraction tomography was developed recently to render refractive index distributions of 100-μm-thick colon tissues with subcellular resolution[35].

Epi-mode tomographic QPI method shows phase reconstruction of cerebral organoids and whole mouse brain with thickness up to 60 μm via deconvolution[36]. Gradient light interference microscopy (GLIM)[37,38] has been developed to suppress multiple scattering via white light, phase-shifting interferometry, which allowed for imaging and analyzing quantitatively opaque structures, such as spheroids and embryos with thicknesses around 300 μm. However, as a widefield technique, GLIM has limited axial resolution and suffers from spatial crosstalk, which mixes diffraction contributions by neighboring points from within the specimen. As a result, the accurate discrimination of cellular boundaries deep within a spheroid remains challenging.

Recent developments in artificial intelligence (AI) and machine learning have brought new opportunities to these challenges. Deep learning enables super-resolution in fluorescence microscopy by training a generative adversarial network to transform confocal microscopy images to stimulated emission depletion images, as well as to transform total internal reflection fluorescence microscopy (TIRF) images to the TIRF-based structured illumination microscopy images[39]. A deep neural network can be trained to virtually refocus a 2D fluorescence image onto 3D volumetric imaging without any axial scanning, additional hardware, or a trade-off of imaging resolution and speed. This framework is also capable of 3D focusing of a single widefield fluorescence image to match confocal microscopy images at different focal planes[40]. An AI-based deep learning algorithm was developed for the automated quantification of the corneal sub-basal nerve plexus for the diagnosis of diabetic neuropathy using corneal confocal microscopy images[41]. AI-assisted adaptive optics methods aim to compensate systematic and tissue-induced aberrations for imaging deep into turbid specimens[42]. Label-free prediction of 3D confocal fluorescence images can be obtained from either transmitted-light microscopy images or electron micrograph inputs[43]. A label-free, volumetric, and automated assessment method has been developed for immunological synapse using optical diffraction tomography and deep learning-based segmentation[44].

In this paper, we report artificial confocal microscopy (ACM), a laser scanning QPI system combined with deep learning algorithms, which renders “synthetic” fluorescence confocal images from unlabeled specimens. First, we developed a laser-scanning QPI system, which is implemented as an upgrade module onto an existing laser scanning confocal microscope (LSM 900, Airyscan 2, Zeiss). We validated the boost in sensitivity and axial resolution of the new system by using standard samples and rigorous comparison with the widefield counterpart. Second, we derived a theoretical model based on the first-order Born approximation, which yields an analytic solution for the spatial frequency coverage of the laser scanning QPI system. These results were validated using experiments to measure the transfer function of the instrument. Third, we trained an artificial neural network on pairs of laser-scanning QPI and fluorescence confocal images from the same field of view. Because the QPI module is attached to the same optical path, generating the training data is straightforward and automated, as the fields of view are intrinsically registered. Fourth, we applied the inference of the “computational” neural network to monolayers of “biological” neural networks and found that the resulting 3D images mimic very well those of the ground truth from the confocal fluorescence images. Using these ACM images, we created binary masks for the contour of the cell and applied them back to the QPI (input) data.

Our results show that the measurements of cell volume and dry mass of ACM vs. confocal agree very well. Fifth, we used the ACM images to perform nuclear segmentation and, thus, cell counting, within hepatocyte spheroids. We also showed that the training performed on spheroids suspended in PBS can transfer to specimens suspended in hydrogel, which promises broad applications in tissue engineering.

Results

The ACM imaging system consists of an existing confocal microscope augmented by a laser scanning GLIM system (LS-GLIM). Figure 1a illustrates the ACM setup, which has three main modules: the laser-scanning confocal microscope (LSM 900, Zeiss), the differential interference contrast (DIC) microscope, and the LS-GLIM module. The LS-GLIM assembly shares the laser source from confocal microscopy (see Methods). The two sheared beams that form the DIC image have their relative phase shift controlled by the liquid crystal variable retarder (LCVR). The LCVR was carefully calibrated to produce accurate phase shifts, as described in SI Note 1. For each $\pi/2$ phase shift, the transmitted light photomultiplier tube (PMT) records the resulting interferogram, as shown in Fig. 1b. The quantitative phase images are generated by the phase-retrieval reconstruction and Hilbert integration algorithms described in the GLIM operation[38]. By sharing the same illumination path, the imaging system registers QPI z-stack images and pairs them with confocal fluorescence frames from the same field of view, which serve, respectively, as input and ground truth data for the deep learning algorithm (Figs. 1b–c). Due to the laser-scanning illumination and PMT detection, the noise level is reduced by a factor of 5 compared to the full-field method (See SI Note 2), thus, the spatial sensitivity of the phase images is improved. The ground truth data, i.e., confocal fluorescence images, provide specificity with high axial resolution and signal-to-noise ratio (SNR). Our goal is to use deep learning to infer the fluorescence confocal images from the LS-GLIM input data and, thus, replicate the confocal advantages on unlabeled specimens.

Multichannel EfficientNet-based U-Nets (E-U-Nets) were trained to translate the 3D phase image stack to the corresponding 3D fluorescent image stack. An E-U-Net comprises a standard U-Net where the encoder is replaced with an EfficientNet[45] (Fig. 2a). The multichannel inputs of an E-U-Net are 3 neighboring quantitative phase images along the z-axis, and the output is the corresponding central fluorescent image slice (see Methods). We chose this three-frame set as input to account for the fact that the axial spread in LS-GLIM data is significantly more pronounced than in the confocal fluorescence data, primarily because the input image is obtained in a transmission geometry without a pinhole, while the output is in reflection with a pinhole or Airyscan detector array. Thus, the neural network “learns” the spread mechanism from the three adjacent images and reverses it to produce a sharp ACM frame.

In SI Note 3, we present a full description of the 3D image formation in LS-GLIM for weak-scattering samples, which starts with the inhomogeneous wave equation and considers scattering under the Born approximation[46]. The expression for the signal collected at the detector has a particularly simple and physical intuitive form,

$$s(\boldsymbol{\rho}) \propto \chi(\boldsymbol{\rho}) \circledv \left[U_d(\boldsymbol{\rho}) U_i^*(\boldsymbol{\rho})\right], \quad (1)$$

where χ is the scattering potential of the specimen, U_d and U_i are the detection and illumination functions, defined as the Fourier transforms of their respective pupil functions, * stands for complex conjugation, and $\circledv$ denotes the 3D convolution in the spatial domain, $\boldsymbol{\rho}$. Thus, the point spread function is given by the product $U_d(\boldsymbol{\rho})U_i^*(\boldsymbol{\rho})$, i.e., it improves with both a tighter illumination focus and a broader detection pupil. Note that equation (1) is restricted to weakly scattering specimens, such as the phase edge used to estimate the LS-GLIM coherent transfer function (see Supplementary Fig. S5, S6). These theoretical predictions are comparable with the experimental measurements for various detection NAs. As LS-GLIM uses elastic scattering and operates in transmission mode, the frequencies in the missing cone region of the coherent transfer function cause inferior axial resolution and sectioning compared to confocal fluorescence microscopy. Hence, we rely on the neural network with confocal fluorescence images as reference.

Figure 2b compares images of a 2 μm microbead under widefield GLIM, LS-GLIM, confocal fluorescence microscopy, and the network inference, i.e., the ACM image. Interestingly, the resulting ACM image is characterized by significantly less axial blur than the LS-GLIM input. As described in the SI Note 2, the sensitivity of LS-GLIM is superior to its widefield counterpart, due to the absence of spatial crosstalk and higher sensitivity photon multiplier detector. However, due to the transmission geometry, they are both inferior to the reflection confocal images in terms of axial sectioning. In contrast, the corresponding network inferences, i.e., the ACM images show significantly improved axial resolution and sectioning. The average Pearson correlation coefficient (PCC) of the microbead is 98% and the Peak signal-to-noise ratio (PSNR) is 46.3. The three adjacent LS-GLIM frames used as network input (see methods for details) contain information about the field Laplacian along z, which governs the inhomogeneous wave equation (Supplemental Note 3) and may explain why this network architecture can produce adequate results in terms of 3D reconstructions.

Next, we applied ACM to imaging neural cultures. We used two common stains to tag the Tau and MAP2 proteins[47] (see Methods) the ratios of which are a popular model for differentiating the long axon from smaller dendrites. The confocal fluorescence images from the two channels represent the ground truth and, as before, the corresponding LS-GLIM images were the input data. The training data contained 20 z-stacks of neurons at 10 days in vitro (DIV 10). The results are summarized in Figs. 3a–l. Our results indicate that the overall 3D renderings of the ground truth and their inferences match very well. The PPC and PSNR of channel Tau are 80% and 26.9 respectively for Fig. 3 (a–l). The PPC and PSNR of channel MAP2 are 91% and 29.1 respectively for Fig. 3 (a–l). Occasionally, we found some discrepancies in the dendrites, which translates into lower correlation values. In Figs. 3f, l, the white arrow points to the axon of the neuron. Interestingly, ACM images reduced the pixel-level noise and confocal stripe artifacts present in the training data. In extended data Fig. 3, we compared the power spectra of the neurons from Fig. 3. (a–l) using GT and ACM images. The 3D frequency coverage of the GT and ACM spectra agree, and both reach the theoretical confocal fluorescence resolution limits. The ACM data allows us to delineate individual cells accurately and measure their volumes. Supplemental Video 1 illustrates this

performance on live neurons that have never been labeled. Visually, it is evident that the ACM provides a much sharper decay of the out-of-focus light, i.e., greater depth sectioning, than the original LS-GLIM. Supplemental Video 2 illustrates the time-lapse performance of ACM on unlabeled, dynamic neurons. Of course, the ACM images do not suffer from bleaching or toxicity, while they maintain chemical specificity through computation. As a result, ACM is suitable for studying live cellular systems non-destructively, over large periods of time.

From the ACM images, we computed binary masks corresponding to the cell contours, which were applied back to the input QPI maps to retrieve individual cell dry mass values. From the cell volume and mass, we also extracted the dry mass density for each cell. Figures 3 (m–o, Volume) are rendered using binarized ACM predicted MAP2, and Figs. 3 (p–r, Dry Mass Density) are calculated from ACM predicted MAP2 segmentation and LS-GLIM phase images. The PPC and PSNR of channel MAP2 are 90% and 32.8 respectively for Fig. 3 (m–r). Figures 3s–t compare the volume and dry mass associated with a single cell measured from confocal MAP2 and ACM MAP2 predictions. The average volume of a single cell is determined by the total volume per field of view (FOV) divided by the number of cell body within the FOV. Our results indicate that the volume and dry mass measurements are well-matched with the ground truth, i.e., there is no significant difference between the two distributions ($p >> 0.05$).

To demonstrate ACM's ability to delineate cellular structures inside turbid spheroids, we imaged hepatocyte spheroids (HepG2) suspended in phosphate-buffered saline (PBS) and generated computational stains associated with the deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) (Fig. 4a). The RNA is localized within the nucleus, with a high concentration in the nucleolus (Fig. 4b). The study of RNA is of high current interest, not only because it plays a crucial role in catalyzing cellular processes, but also because it can be used by various viruses to encode their genetic information[48]. The two ground truth stains, 7-aminoactinomycin D (7-ADD) and SYTO RNASelect Green (see Methods), and their associated inferences, enable us to generate semantic segmentations and annotate the spheroid into "nuclei" and "nucleoli", respectively. The entire "spheroid" represents our third class and is obtained as the non-background regions in the LS-GLIM data. As shown in Fig. 4c and SI Note 4, the actual and imputed fluorescent maps show good agreement. The PPC and PSNR of channel DNA are 84% and 24.9, respectively. As detailed in Methods, we apply a threshold on the ACM and phase image to generate 3D semantic segmentation maps, which we use to measure the dimension of the spheroid. The intersection of the RNA and DNA labels provides the annotation for the nucleoli. Our results show that, across the 20 spheroids studied in this work, the total nuclear mass is proportional to the spheroid mass. This dependence is shown in Fig. 4d, where the slope of the linear regression (0.42) indicates that about 42% of the spheroid mass is contributed by the nuclei. These results agree well with the same procedure applied to the ground truth confocal images.

Automatic instance segmentation of cells inside spheroids was performed by 3D marker-controlled watershed on the estimated DNA signal. The markers were determined through a 2D Hough voting on each slice in the z-stack basis (Fig. 5). The result of the Hough voting is a volume with a unique marker on the spheroid, which resembles a column tracking

the center of the nucleus through the focus (see Methods for details). The result of the watershed is a 3D volume with a unique label for each nucleus within the spheroid, which enables the calculation of parameters on individual cells. In order to compare the mass, volume, and mass density distributions, we computed the relative spread, σ/μ, with σ the standard deviation and μ the mean associated with the best Gaussian fit. Our data indicate that nuclear density (Fig. 5d, $\sigma/\mu = 0.2$) has a significantly narrower distribution compared to the distributions of nuclear mass (Fig. 5b, $\sigma/\mu = 0.9$) and volume (Fig. 5c, $\sigma/\mu = 0.8$). These observations indicate that the dry mass density is a much more uniform parameter across different cells. Given the broad distribution of volumes and masses, this result shows that a change in volume is accompanied by an almost linear change in mass.

Discussion:

The principles of AI and confocal microscopy were both formulated in the mid-1950s. Since then, the two technologies have taken on independent trajectories, with confocal leading to an entirely new class of scanning imaging modalities and AI giving rise to a variety of applications, from digital assistants to autonomous vehicles. Furthermore, in the past several years, it has become apparent that AI algorithms are valuable tools for extracting knowledge from optical images. As such, the two fields are intersecting again, and this combination appears to hold exciting prospects for biomedicine.

We developed artificial confocal microscopy to combine the benefits of non-destructive imaging from QPI with the depth sectioning and chemical specificity associated with confocal fluorescence microscopy. Augmenting an existing laser scanning microscope with a QPI module (LS-GLIM) we can easily collect pairs of registered images from the phase (input data) and fluorescence (ground truth) channels, which allow us to generate colocalized ground truth – input pairs of images. As expected, the transmission quantitative phase image exhibits a much stronger elongation along the z-axis, as the scattering wavevector (or momentum transfer) has a much shorter z-component than in the reflection geometry. These pairs of images are used to train a neural network (Efficient U-Net) to perform image-to-image translation from the LS-GLIM to the confocal fluorescence signal. The final ACM image presents the characteristics of the confocal image, with good axial sectioning and chemical specificity (see Figs. 2–3). Applying ACM to unlabeled cells allows us to nondestructively translate the confocal microscopy features to dynamic imaging (see Supplemental Videos 1–2). The image formation for weak-scattering samples in LS-GLIM was described here for the first time, to our knowledge. The theoretical model agrees with the experimental data on the system's transfer function by imaging a phase edge.

The confocal geometry, advances the QPI imaging capability in several ways, as follows. First, by illuminating one point at a time, the confocal eliminates the spatial crosstalk that affects widefield methods. With this illumination, the noise from the neighboring points is lowered. Second, the PMT array provides a much more sensitive detection, which, together with the first point above, yields the overall higher sensitivity, both spatially and temporally, that we capture in our data. Third, the backscattering geometry yields higher axial spatial frequency coverage, resulting in stronger sectioning. This quality is obvious in our data, both ground truth and inference.

By overcoming the spatial crosstalk limitations associated with widefield methods, ACM has the potential to provide new data for studying turbid cellular systems. Measuring quantitatively functional parameters from organoids and spheroids can be useful in a variety of applications of biological and clinical relevance. Using the artificial fluorescence images generated by the neural network, we segmented individual nuclei within the 3D structures, which can be used not just for cell counting but also for computing individual nuclear volumes. Furthermore, by creating annotations from the ACM images and applying them back to the input phase images, we extracted dry mass information from individual nuclei, independently from the nuclear volume. Our results indicate that, on average, 42% of the spheroid mass is contained in the nuclei. We also found that the nuclear dry mass density distribution is significantly narrower than the volume and mass counterparts. Nevertheless, ACM images can differ from confocal fluorescence images. The potential source of error could come from the training corpus including the confocal fluorescence images and LS-GLIM phase images. Specifically, LS-GLIM phase images lack sensitivity to spatial frequencies within the missing cone region, while the confocal fluorescence images are affected by variations in the staining level. Moreover, the contrast in LS-GLIM images comes from the intrinsic inhomogeneity of refractive index distributions in tissues, resulting in much more details compared to the confocal fluorescence counterparts.

Finally, we demonstrated that the network training can be transferred between spheroids suspended in different media with no additional training, which provides versatility to our technique. However, the accuracy of prediction is lower than those of spheroids in PBS, which is expected since the shape of the spheroids in PBS and hydrogel is quite different, and the network is never trained on those data. We anticipate that ACM can be potentially adopted at a broad scale because the LS-GLIM module can be readily added to any existing laser confocal system, while the data for training can be acquired with ease. ACM provides complementary information to that from other laser-scanning techniques, as the acquisition is not limited by photobleaching and toxicity, while the axial resolution is maintained at confocal levels.

Methods

ACM system

The experimental set-up for ACM is a multichannel imaging system, which consists of confocal microscopy (LSM 900, Zeiss) and laser-scanning gradient light interference microscopy (LS-GLIM). The LS-GLIM module upgrades a laser scanning confocal microscope outfitted with DIC optics by providing phase-shifting assembly capability (Fig. 1(a)). The laser-scanning interference microscope shares the same two-laser lines (488nm, 561nm) of the confocal microscope. The laser source from the confocal microscopy goes up through the matched DIC prism and objective (63×, 40×) and then is scattered by the sample. After the sample, the light is collected by the condenser of the DIC microscope. Next, the light travels through the phase-shifting assembly, which consists of a liquid crystal variable retarder (LCVR, Thorlabs) followed by a linear polarizer. In order to allow the liquid crystal to modulate the phase shift between the two orthogonal polarizations, we removed the analyzer that normally sits inside the condenser. The stabilization time of the

LCVR is approximately 70 ms. Four intensity frames are recorded by the photomultiplier tube (PMT, Zeiss) corresponding to each $\pi/2$ phase shift, as shown in Fig. 1(b). The acquisition time of each frame is approximately the same as for a confocal fluorescence image, which depends on the dwell time and pixel numbers set for the image acquisition. The dwell time for all the images was chosen to be 1.2 μs, such that for an image with 1744×1744 pixels, the acquisition time is ~ 3.7s. The quantitative phase images are generated in real-time by the phase-retrieval reconstruction algorithm and Hilbert transform algorithm[38]. The system registers pairs of z-stack images from both the confocal fluorescence and quantitative phase, which serve, respectively, as ground truth and input images for machine learning (Fig. 1(b)). The z-sampling was chosen to be 0.2 μm, 0.2 μm, and 1 μm for microbeads, neurons, and spheroids, respectively. The x-y sampling was 0.09 μm for all the data presented in this paper.

Network training: Multichannel EfficientNet-based U-Nets (E-U-Nets) were trained with paired phase and fluorescent images. The input channels of an E-U-Net are three neighboring phase slices, and the output is the corresponding central fluorescent slice. This network design allows an E-U-Net to use information from phase images acquired at multiple neighboring imaging planes to better predict the fluorescent image.

The network architecture of a multichannel E-U-Net is shown in Fig. 2a and Fig. S7. It represents a modification of a standard U-Net where the encoder is replaced with an EfficientNet[45]. The EfficientNet generally has a powerful capacity for feature extraction but is relatively small in network size. Training an E-U-Net from scratch can be challenging when the number of paired phase and fluorescent images is limited. To mitigate this challenge, a transfer learning strategy was used in the E-U-Net training. Specifically, the weights of the EfficientNet encoder were initialized with weights pre-trained on an ImageNet dataset[49] for an image classification task. The ImageNet is a benchmark image set that contains millions of labeled nature images.

In this study, a neuron dataset, a spheroid cell dataset, and a bead dataset were used for training, validating, and testing the E-U-Nets. The neuron dataset contained 22 image stacks that each contained 300 neuron phase images of size 1744×1744 pixels and their related two-channel fluorescent images, which correspond to fluorescent signals from Tau and MAP2 proteins, respectively. The spheroid cell dataset contained 21 stacks that each contained 100 spheroid cell phase images of size 1744×1744 pixels and the related two-channel fluorescent images, which correspond to fluorescent signals from DNA and RNA, respectively. The bead dataset contained 18 image stacks that each contained 250 bead phase images of size 128×128 pixels and the associated fluorescent images. To facilitate network training, the pixel values in each fluorescent image stack were scaled to a range of [0, 255.0]. This was accomplished as: $x_o = 255.0 \times \frac{x_i - x_{0.01\%}}{x_{99.99\%} - x_{0.01\%}}$, where $x_{0.01\%}$ and $x_{99.99\%}$ represent the 0.01%-th and 99.99%-th values among all the pixel values in the image stack after they were sorted in non-decreasing order; x_i and x_o represent the original value and the scaled value of a pixel, respectively. The estimated fluorescent image stack was subsequently rescaled to its original range using $x_i = \frac{x_0}{255.0}(x_{99.99\%} - x_{0.01\%}) + x_{0.01\%}$. For those

image stacks without ground truth values, the $\hat{x}_{0.01\%}$ and $\hat{x}_{0.99\%}$ can be estimated as the average of $x_{0.01\%}$ and $x_{99.99\%}$ related to the ground truth values in the training set.

Considering the limited number of image stacks in the three datasets described above, a 3-fold cross-validation (CV) approach was employed to train and validate the E-U-Nets after a few testing image stacks were held out for the E-U-Net testing. For a given dataset in which the testing stacks have been held out, the 3-fold CV approach involves randomly dividing all the stacks in the dataset into 3 folds of approximately equal size. The first two folds and the remaining one-fold were treated as a training set and a validation set to train and validate E-U-Nets, respectively. The procedure was repeated three times; each time, a different fold was treated as the validation set. The three procedures resulted in the validation of the E-U-Nets on each image stack. The trained E-U-Nets were finally tested on the held-out unseen testing samples. Details related to the cross-validation of E-U-Nets on the neuron, spheroids cell, and bead datasets are described below.

For the neuron dataset, two separate E-U-Nets were trained: one to translate phase images into each of the two-channel fluorescent images. The EfficientNet-B7 network was employed in the two E-U-Nets. The network architecture of the EfficientNet-B7 is shown in Fig. S7. Two neuron image stacks were held out as unseen testing data; the remaining 20 stacks were employed in the 3-fold CV process described above. In the 3-fold CV process, the 20 image stacks were randomly divided into three folds that contained 6, 7, and 7 image stacks, respectively. For each data split, the E-U-Nets were trained by minimizing a mean square error (MSE) loss function that measures the difference between the predicted fluorescent images and their corresponding ground truth values. The loss function was minimized by the use of an ADAM optimizer[50] with a learning rate of 5×10^{-4}, which was empirically determined. In each training iteration, a batch of paired 3 neighboring phase images and the corresponding central fluorescent image were sampled from the training image stacks, and then randomly cropped into patches of 512×512 pixels as training samples to train the networks. The batch size was set to 4. A decaying strategy was applied to the learning rate to mitigate the overfitting by multiplying the learning rate by 0.8 when the validation MSE loss did not decrease for consecutive 10 epochs. An epoch is a sequence of iterations that walk through all the image slices in the training set. The validation MSE loss was computed between the predicted fluorescent images and their ground truth values for validation images. In the network training, an early-stopping strategy was employed to determine the end of the network training. Specifically, at the end of each epoch, the being-trained E-U-Net model was evaluated by computing the average of Pearson correlation coefficients (PCCs) between the predicted fluorescent images and the related ground truth values. The network training stopped if the average validation PCC did not increase for 20 epochs as shown in Fig. S8. The two figures show the average training and validation stopping rule metric for training the two E-U-Nets respectively in one of the three training procedures of the 3-fold CV process. After the E-U-Nets were trained, the performances of the trained networks were evaluated on the validation set by computing the peak signal-to-noise ratio (PSNR) and PCC between the predicted fluorescent stacks and the related ground truth values. The 3-fold CV process resulted in validation results for each of the 20 stacks. These validation results were combined and are reported in SI Note 4. In

addition, the E-U-Nets trained in the CV process were tested in the two unseen stacks. The corresponding Pearson correlation coefficients (PCCs) and PSNR are presented in SI Note 4.

For the spheroids cell dataset, two separate E-U-Nets were trained for each fluorescent channel. The EfficientNet-B7 network was employed as the encoder in the two trained E-U-Nets. Two spheroid cell image stacks were held out for E-U-Net testing; the remaining 19 stacks were randomly split into three folds that contain 6, 6, and 7 stacks, respectively, in the 3-fold CV process. The other training settings were the same as those described above for network training on the neuron dataset. The training and validation PCCs over epochs correspond to training the two E-U-Nets in one of the three training procedures of the 3-fold CV process are displayed in Fig. S9. The 3-fold CV results related to PSNR and PCC are reported in SI Note 4. The results tested on two unseen testing stacks are shown in SI Note 4.

For the bead dataset, a single E-U-Net was built for the phase-to-fluorescent image translation. The EfficientNet-B0 was employed as the encoder in the E-U-Net. The architecture of the EfficientNet-B0 network is shown in Fig. S7. One of the bead image stacks was held out as an unseen testing stack for the E-U-Net testing; the remaining 17 bead stacks were randomly divided into three folds that each contains 5, 6, and 6 image stacks, respectively, for the 3-fold CV process. Paired images of size 128×128 pixels were employed for the E-U-Net training. The batch size was 32. The other training settings were the same as those for the network training on neuron and spheroid cell datasets, as described above. The training and validation stopping rule metric over epochs for one of the three training procedures of the 3-fold CV process are displayed in Fig. S10. The 3-fold CV results related to PSNR and PCC performances are reported in SI Note 4. The results on the unseen bead stack are shown in SI Note 4.

The E-U-Nets were implemented by use of the Python programming language with libraries including Python 3.6 and TensorFlow 1.14. The model training, validation, and testing were performed on an NVIDIA Tesla V100-GPU with 32 GB VRAM. The E-U-Net training on the neuron dataset and spheroid dataset took approximately 24 hours. The E-U-Net training on the bead dataset took approximately 2 hours. The inference time for a fluorescent image slice of 1744×1744 pixels was approximately 400 *ms*.

Neuron analysis: The volume of neurons was calculated from the ACM images using binary masks with background thresholding. The 3D dry mass distribution was generated with the multiplication of binary masks and the 3D dry mass distribution from the QPI images[51]. The 3D dry mass density is linearly related to the depth-resolved phase maps as

$$M(x, y, z) = \frac{\lambda}{2\pi\gamma\delta z}\phi(x, y, z), \quad [2]$$

where λ is the wavelength of the illumination and $\gamma \simeq 0.2$, the refractive increment, which lies within the range of 0.18–0.21 ml/g for most biological samples[52]. δz is the z sampling, which for our LS-GLIM is ~ 1 μm. $\phi(x, y, z)$ is the measured phase image at each z-plane.

Spheroid analysis: Three-dimensional semantic segmentation maps were generated from the estimated fluorescent signals corresponding to the RNASelect and 7-ADD stains by applying fixed thresholds for the entire data. This map of RNA and DNA-stained regions was further refined by assigning a “nucleoli” label to the RNA inside the DNA regions. To generate a map labeling the “spheroid”, a threshold was applied to the quantitative phase signal after Hilbert demodulation[51]. Fields of view were acquired to contain a single spheroid, and phase values coincident with the assigned label (“nucleus”, “spheroid”) were totaled on a per-spheroid basis to report on the dry mass and volume.

Automated 3D cell counting: To segment our images into individual nuclei, we used a 3D variation of the marker-controlled watershed on the estimated DNA images[53]. We note that the ACM data lacked the unwanted pixel-level noise typically associated with photon-starved fluorescent images. This technique requires the image to be annotated into sample and background regions with a non-overlapping marker used to identify the cell. We performed 2D Hough voting which is used to identify the center of the nucleus in each z-slice, producing what resembles a curve through the z-dimension. To regularize our approach, we applied a 3×3 blur to correct for minor disconnects in our segmentation algorithm. The result of our watershed approach is a 3D volume with a unique label annotating each nucleus (Fig. 5a). To validate our method, we compared our results to a manual cell count performed in AMIRA. We obtained 142 cells counted automatically vs 136 cells counted manually (4 % error). The principal disagreement was due to undercounting touching cells. This procedure was implemented in MATLAB using the “imfindcirlces” and “watershed” commands.

Sample preparation:

Hippocampal neuron preparation: All procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Illinois Urbana-Champaign and conducted per the guidelines of the U.S National Institute of Health (NIH). For our neuron imaging experiments, we used primary hippocampal neurons harvested from dissected hippocampi of Sprague-Dawley rat embryos at embryonic day 18. Dissociated hippocampal neurons were plated on multi-well plates (Cellvis, P06-20-1.5-N) that were pre-coated with poly-D-lysine (0.1 mg/ml; Sigma-Aldrich). Hippocampal neurons were incubated for three hours under the condition of 37°C and 5% CO_2 in a plating medium containing 86.55% MEM Eagle’s with Earle’s BSS (Lonza), 10% Fetal Bovine Serum (re-filtered, heat-inactivated; ThermoFisher), 0.45% of 20% (wt./vol.) glucose, 1× 100 mM sodium pyruvate (100x; Sigma-Aldrich), 1× 200 mM glutamine (100x; Sigma-Aldrich), and 1x Penicillin/ Streptomycin (100x; Sigma-Aldrich) in order to help attachment of neurons (300 cells/mm^2). The plating media was aspirated and replaced with maintenance media containing Neurobasal growth medium supplemented with B-27 (Invitrogen), 1% 200 mM glutamine (Invitrogen), and 1% penicillin/streptomycin (Invitrogen) and incubated for 10 days at 37 °C, in the presence of 5% CO_2. Hippocampal neurons were maintained for 14 days before performing immunostaining.

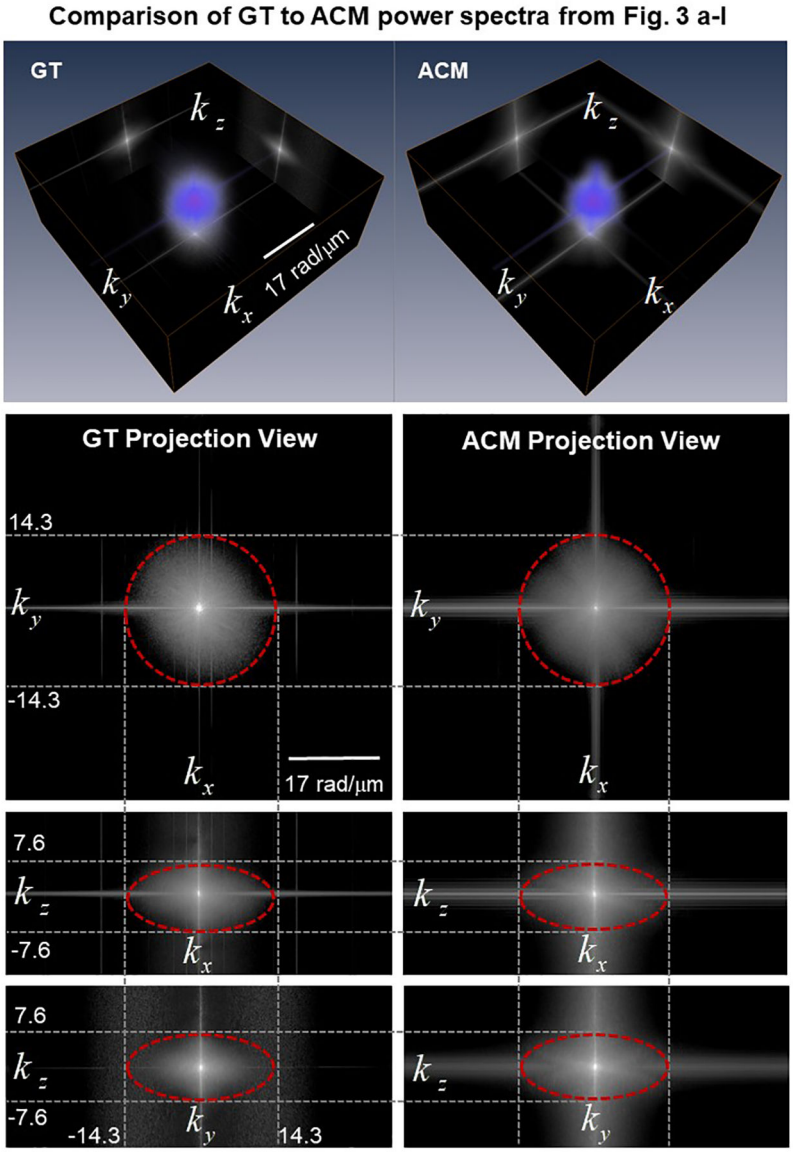
Immunostaining protocol: Neurons were stained with antibodies for Tau (Abcam, ab80579) and MAP2 (Abcam ab32454) to localize axons and dendrites. Neurons were fixed

with freshly prepared 4% paraformaldehyde (PFA) for 15 minutes following 0.5% Triton-X for 10 minutes and 2% bovine serum albumin (BSA, ThermoFisher) for 2 hours incubation in 4°C. Hippocampal neurons were incubated for 8 hours at 4°C with anti-Tau antibodies that were diluted to 1:250 in 5% BSA. After washing with PBS, neurons were exposed for 8 hours at 4°C to goat anti-mouse secondary antibody (Abcam, ab205719) which was diluted to 1:500 in 5% BSA. Hippocampal neurons were then incubated in anti-MAP2 antibody (1:500 dilution) in 5% BSA for 8 hours, followed by goat antirabbit secondary antibody (Abcam, ab205718, 1:1000 dilution) in 5% BSA for 8 hours at 4°C.

Liver cancer spheroid (HepG2 cells): Human hepatocarcinoma cells (HepG2, ATCC) were cultured in T-75 flasks with DMEM (Thermo) + 10% fetal bovine serum + 1% penicillin-streptomycin (Gibco) for 7 days and formed spontaneous pre-formed spheroids. The flasks were incubated at 37C and 5% CO^2. Media was replaced every 2–3 days. Spheroids were incubated with TrypLE Express (Thermo) for 10 minutes to detach pre-formed spheroids of approximately 100~200 μm in diameter from the culture flask. The passage number used was between 2–6.

Pre-formed spheroids were plated on poly-d-lysine coated glass-bottom dishes. The spheroids were incubated for 10 minutes to allow for attachment. Then, they were covered with a collagen hydrogel (bovine collagen type 1, Advanced Biomatrix). The cells were incubated for 3 days to allow for cellular reorganization into a regular spheroidal shape. The spheroids were first fixed in a 1:1 ratio of methanol: acetone at −4°C for 20 minutes. Cells fixed using this method do not need an additional permeabilization step due to the acetone. The cell nucleus was stained using 7-aminoactinomycin D (7-AAD red, 6163, ThermoFisher) by adding 1μL of the stock stain into 1mL of PBS. The cell RNA was stained using SYTO RNASelect Green (S32703, ThermoFisher) by first creating a 5μM working solution and then adding 100μL of the working solution to 900μL of PBS. The samples were stained at room temperature for 30 minutes before rinsing once. Two kinds of samples in PBS or hydrogel were imaged after staining.

Extended Data



Extended Data Fig. 3. Comparison of GT to ACM power spectra from Fig. 3 a–l.

Comparison of GT to ACM power spectra from Fig. 3 a–l. Contours circumscribing theoretical resolution limits of confocal fluorescence system (GT) are shown in as red dotted circles. The theoretical lateral resolution of the system is 0.22 μm (NA=1.3, 1 Airy Unit (AU), excitation wavelength at 561 nm), corresponding to a maximum lateral frequency of 14.3 rad/ μm. The theoretical axial resolution of the system is about 0.50 μm, corresponding to a maximum axial frequency of 6.3 rad/ μm. The 3D frequency coverage of the GT and ACM spectra agree, and both reach the theoretical resolution limits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

This work is supported by the National Science Foundation (CBET0939511 STC, NRT-UtB 1735252, CBET-1932192), the National Institute of General Medical Sciences (GM129709), and the National Cancer Institute (CA238191).

Data availability:

Due to the size consideration, the data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Freedman BS et al. Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. Nature communications 6, 1–13 (2015).
2. Langhans SA Three-dimensional in vitro cell culture models in drug discovery and drug repositioning. Frontiers in pharmacology 9, 6 (2018). [PubMed: 29410625]
3. Bissell MJ, Hall HG & Parry G How does the extracellular matrix direct gene expression? Journal of theoretical biology 99, 31–68 (1982). [PubMed: 6892044]
4. Laschke MW & Menger MD Life is 3D: boosting spheroid function for tissue engineering. Trends in biotechnology 35, 133–144 (2017). [PubMed: 27634310]
5. Bell CC et al. Comparison of hepatic 2D sandwich cultures and 3D spheroids for long-term toxicity applications: a multicenter study. Toxicological Sciences 162, 655–666 (2018). [PubMed: 29329425]
6. Fong ELS, Toh TB, Yu H & Chow EK-H 3D culture as a clinically relevant model for personalized medicine. SLAS TECHNOLOGY: Translating Life Sciences Innovation 22, 245–253 (2017).
7. Kamm RD et al. Perspective: The promise of multi-cellular engineered living systems. APL Bioengineering 2, 040901 (2018). [PubMed: 31069321]
8. Cvetkovic C et al. Three-dimensionally printed biological machines powered by skeletal muscle. Proceedings of the National Academy of Sciences 111, 10125–10130 (2014).
9. Williams BJ, Anand SV, Rajagopalan J & Saif MTA A self-propelled biohybrid swimmer at low Reynolds number. Nature communications 5, 1–8 (2014).
10. Chen X & Korotkova O Optical beam propagation in soft anisotropic biological tissues. Osa Continuum 1, 1055–1067 (2018).
11. Tuchin VV & Society of Photo-optical Instrumentation Engineers. Tissue optics : light scattering methods and instruments for medical diagnosis. 2nd edn, (SPIE/International Society for Optical Engineering, 2007).
12. Chen W et al. High-throughput image analysis of tumor spheroids: a user-friendly software application to measure the size of spheroids automatically and accurately. J Vis Exp, 51639, doi:10.3791/51639 (2014). [PubMed: 25046278]
13. Minsky MS Memoir on Inventing the Confocal Scanning Microscope. Scanning 10, 128 (1988).
14. Wilson T & Sheppard C Theory and practice of scanning optical microscopy. (Academic Press, 1984).
15. Diaspro A Optical Fluorescence Microscopy. Optical Fluorescence Microscopy by Alberto Diaspro. Berlin: Springer, 2011. ISBN: 978-3-642-15174-3 1 (2011).
16. Lippincott-Schwartz J, Altan-Bonnet N & Patterson GH Photobleaching and photoactivation: following protein dynamics in living cells. Nature Cell Biology, S7 (2003).
17. Hoebe RA et al. Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. Nature biotechnology 25, 249–253 (2007).
18. Crivat G & Taraska JW Imaging proteins inside cells with fluorescent tags. Trends in biotechnology 30, 8–16 (2012). [PubMed: 21924508]
19. Graf BW & Boppart SA Imaging and analysis of three-dimensional cell culture models. Live Cell Imaging: Methods and Protocols, 211–227 (2010).

20. North AJ Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. The Journal of cell biology 172, 9–18 (2006). [PubMed: 16390995]
21. Hoover EE & Squier JA Advances in multiphoton microscopy technology. Nature photonics 7, 93–101 (2013). [PubMed: 24307915]
22. Stelzer EH et al. Light sheet fluorescence microscopy. Nature Reviews Methods Primers 1, 1–25 (2021).
23. Choi WJ, Pepple KL & Wang RK Automated three-dimensional cell counting method for grading uveitis of rodent eye in vivo with optical coherence tomography. Journal of biophotonics 11, e201800140 (2018). [PubMed: 29797544]
24. Huang Y et al. Optical coherence tomography detects necrotic regions and volumetrically quantifies multicellular tumor spheroids. Cancer research 77, 6011–6020 (2017). [PubMed: 28904062]
25. Schnell M et al. High-resolution label-free imaging of tissue morphology with confocal phase microscopy. Optica 7, 1173–1180 (2020).
26. Hase E et al. Scan-less confocal phase imaging based on dual-comb microscopy. Optica 5, 634–643 (2018).
27. Singh VR et al. Studying nucleic envelope and plasma membrane mechanics of eukaryotic cells using confocal reflectance interferometric microscopy. Nature communications 10, 1–8 (2019).
28. Liu C et al. High-speed line-field confocal holographic microscope for quantitative phase imaging. Optics express 24, 9251–9265 (2016). [PubMed: 27137541]
29. Popescu G Quantitative phase imaging of cells and tissues. (McGraw-Hill, 2011).
30. Park Y, Depeursinge C & Popescu G Quantitative phase imaging in biomedicine. Nat Photonics 12, 578, doi:10.1038/s41566-018-0253-x (2018).
31. Chen X, Kandel ME & Popescu G Spatial light interference microscopy: principle and applications to biomedicine. Advances in Optics and Photonics 13, 353–425 (2021). [PubMed: 35494404]
32. Chen X, Kandel ME, Hu C, Lee YJ & Popescu G Wolf phase tomography (WPT) of transparent structures using partially coherent illumination. Light: Science & Applications 9, 1–9 (2020).
33. Merola F et al. Tomographic flow cytometry by digital holography. Light: Science & Applications 6, e16241–e16241, doi:10.1038/lsa.2016.241 (2017).
34. Chen M, Ren D, Liu H-Y, Chowdhury S & Waller L Multi-layer Born multiple-scattering model for 3D phase microscopy. Optica 7, 394–403 (2020).
35. Hugonnet H et al. Multiscale label-free volumetric holographic histopathology of thick-tissue slides with subcellular resolution. Advanced Photonics 3, 026004 (2021).
36. Ledwig P & Robles FE Epi-mode tomographic quantitative phase imaging in thick scattering samples. Biomedical optics express 10, 3605–3621 (2019). [PubMed: 31360607]
37. Kandel ME et al. Epi-illumination gradient light interference microscopy for imaging opaque structures. Nature Communications 10, 4691, doi:10.1038/s41467-019-12634-3 (2019).
38. Nguyen TH, Kandel ME, Rubessa M, Wheeler MB & Popescu G Gradient light interference microscopy for 3D imaging of unlabeled specimens. Nat Commun 8, 210, doi:10.1038/s41467-017-00190-7 (2017). [PubMed: 28785013]
39. Wang H et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature methods 16, 103–110 (2019). [PubMed: 30559434]
40. Wu Y et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. Nature methods 16, 1323–1331 (2019). [PubMed: 31686039]
41. Williams BM et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. Diabetologia 63, 419–430 (2020). [PubMed: 31720728]
42. Cheng S, Li H, Luo Y, Zheng Y & Lai P Artificial intelligence-assisted light control and computational imaging through scattering media. Journal of innovative optical health sciences 12, 1930006 (2019).
43. Ounkomol C, Seshamani S, Maleckar MM, Collman F & Johnson GR Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. Nature methods 15, 917–920 (2018). [PubMed: 30224672]

44. Lee M et al. Deep-learning-based three-dimensional label-free tracking and analysis of immunological synapses of CAR-T cells. Elife 9, e49023 (2020). [PubMed: 33331817]
45. Tan M & Le Q in International Conference on Machine Learning. 6105–6114 (PMLR).
46. Chen X & Korotkova O Probability density functions of instantaneous Stokes parameters on weak scattering. Optics Communications 400, 1–8 (2017).
47. Kanai Y & Hirokawa N Sorting mechanisms of tau and MAP2 in neurons: suppressed axonal transit of MAP2 and locally regulated microtubule binding. Neuron 14, 421–432 (1995). [PubMed: 7857650]
48. Rein A Retroviral RNA packaging: a review. Positive-Strand RNA Viruses, 513–522 (1994).
49. Kingma DP & Ba J Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
50. Deng J et al. in 2009 IEEE conference on computer vision and pattern recognition. 248–255 (Ieee).
51. Kandel ME et al. Multiscale Assay of Unlabeled Neurite Dynamics Using Phase Imaging with Computational Specificity. ACS Sensors 6, 1864–1874, doi:10.1021/acssensors.1c00100 (2021). [PubMed: 33882232]
52. Barer R Determination of dry mass, thickness, solid and water concentration in living cells. Nature 172, 1097–1098 (1953). [PubMed: 13111263]
53. Gonzalez RC & Woods RE (Prentice hall Upper Saddle River, NJ, 2002).

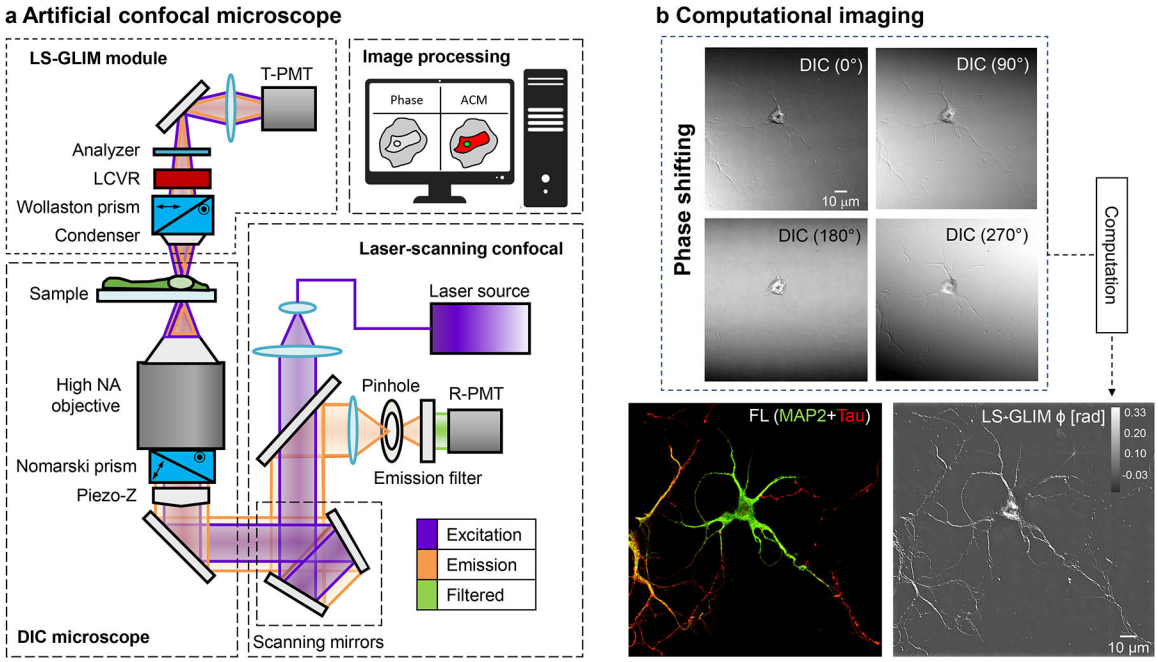


Figure 1.

a, The ACM system consists of a laser-scanning confocal assembly, differential interference contrast (DIC) microscope, and LS-GLIM module. The quantitative phase imaging was conducted with the green laser line (488 nm) of the confocal excitation. The interferogram was recorded at each point in the scan by the transmission-PMT (T-PMT). The fluorescence images were captured by the reflection-PMT of the confocal module. **b**, Four phase-shifting frames are recorded and used to reconstruct the quantitative phase image. The confocal fluorescence image (FL) serves as the ground truth, and the phase image (LS-GLIM) is the input for the network training.

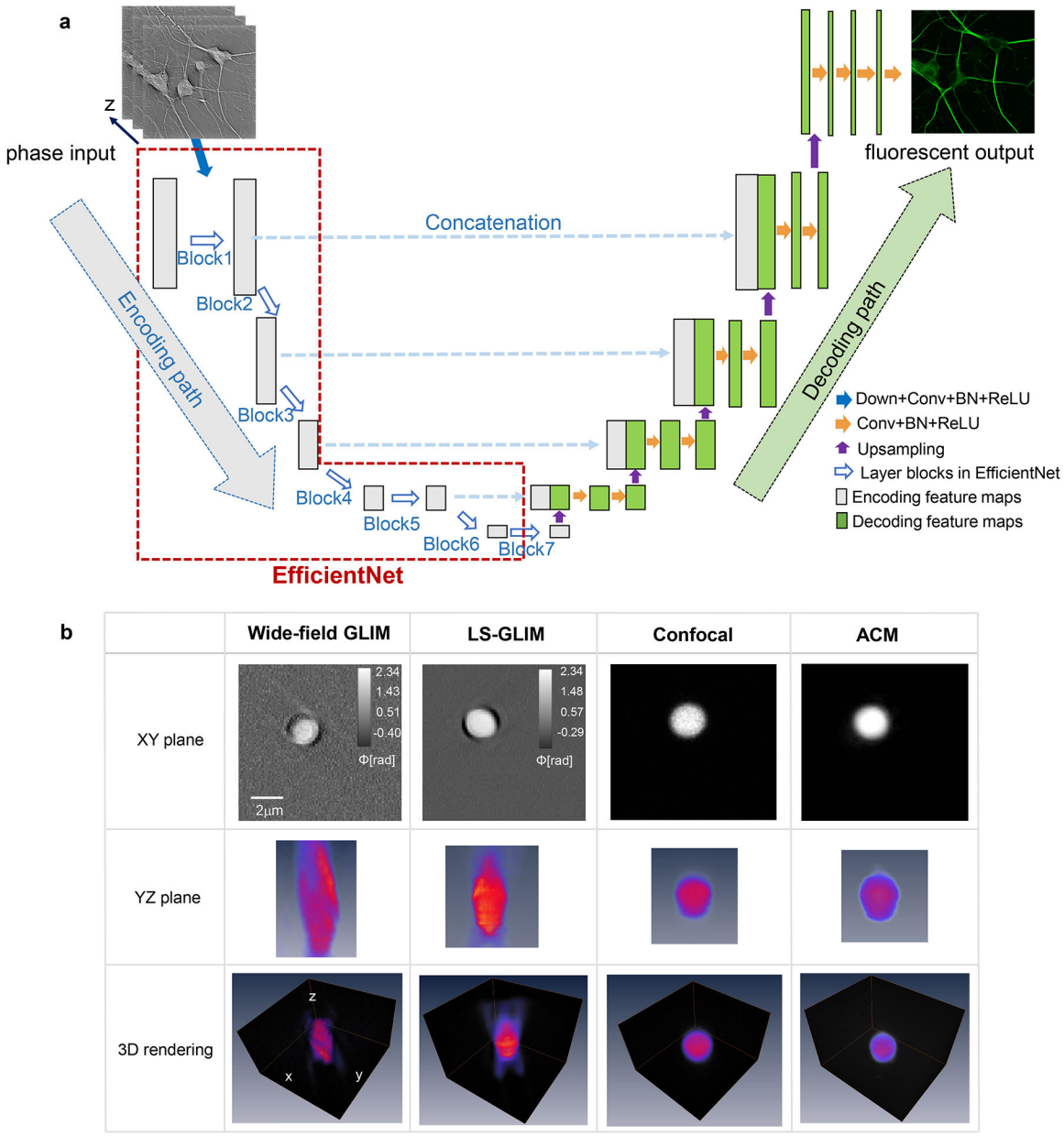


Figure 2.

a, Network architecture for translating phase images into confocal fluorescence signals. It is a U-Net variant that uses an EfficientNet as the encoder. The input of the Efficient-Unet consists of 3 adjacent quantitative phase images along the z-axis, and its output is the corresponding middle fluorescent slice. **b**, Comparison of 2 μm bead (63x/1.3) tomograms in widefield GLIM, LS-GLIM, confocal, and ACM, as indicated. The elongation of the beads in widefield and LS-GLIM is due to the missing frequencies in the transmission geometry. On the other hand, the predicted ACM images replicate the confocal sectioning and resolution.

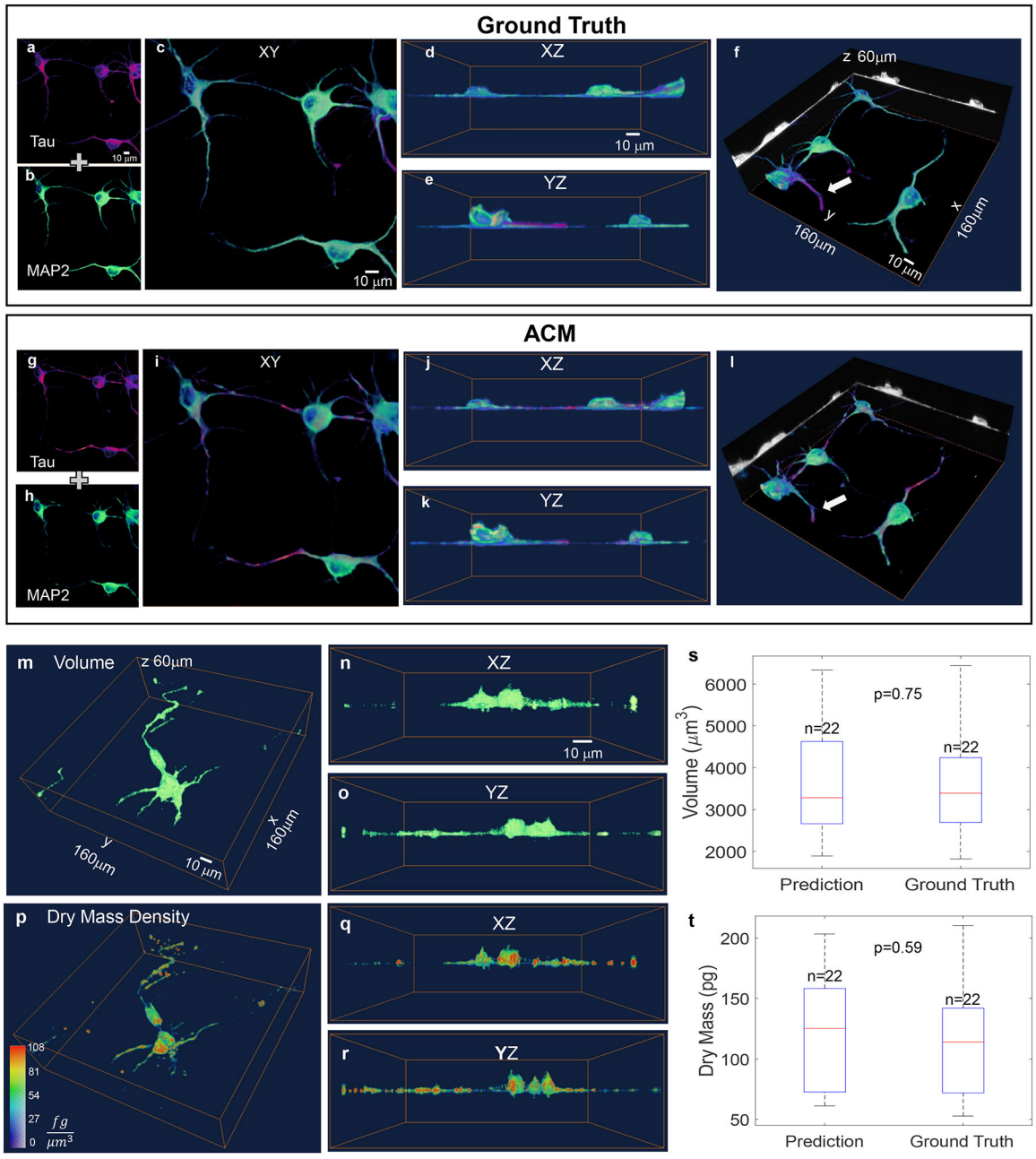


Figure 3.

2D comparison of ground truth from confocal fluorescence (**a**, Tau, **b**, MAP2) and predicted fluorescence (**g**, Tau, **h**, MAP2). 3D comparison of ground truth from confocal fluorescence (**c**, XY, **d**, XZ, **e**, YZ, **f**, 3D tomogram) and predicted fluorescence (**i**, XY, **j**, XZ, **k**, YZ, **l**, 3D tomogram). Volume from binarized ACM predicted MAP2 (**m**, 3D rendering, **n**, XZ **o**, YZ). Dry mass density distribution (**p**, 3D dry mass density **q**, XZ **r**, YZ) based on ACM predicted MAP2 segmentation and LS-GLIM phase images. The box plots of volume (**s**) and dry mass calculation (**t**) for a single cell from confocal MAP2 (Ground Truth) and ACM predictions (Prediction) for 10 days *in vitro* (DIV) neurons. Each box plot shows the median

(red), 25th and 75th percentiles of the sample (the bottom and top of each box), and range (whisker). *P*-values are of the unpaired two-sided t-tests.

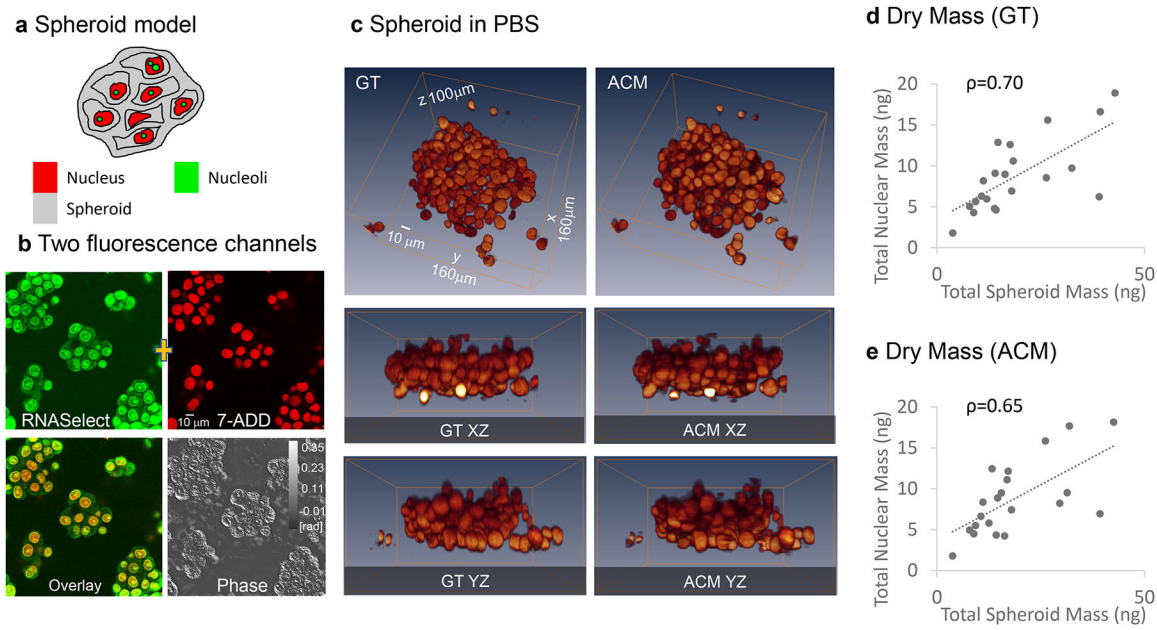


Figure 4.

Label-free intracellular segmentation in turbid spheroids. **a & b**, Cellular compartments were stained using RNA and DNA sensitives stains. DNA is used to identify the nucleus and dense concentrations of RNA inside the nucleus are associated with nucleoli. **c**, 3D comparison of GT and ACM predicted tomography of a spheroid (40x/1.3). For all 20 spheroids, we calculated the nuclear dry mass and volumes generated from the imputed signal. **d**, **e**, Total nuclear dry-mass tracks closely with total spheroid mass, Person correlation coefficient ρ=+0.65 (the slope of the linear fit is 0.42). These results agree well with the same procedure applied to the ground truth confocal images **(d)**.

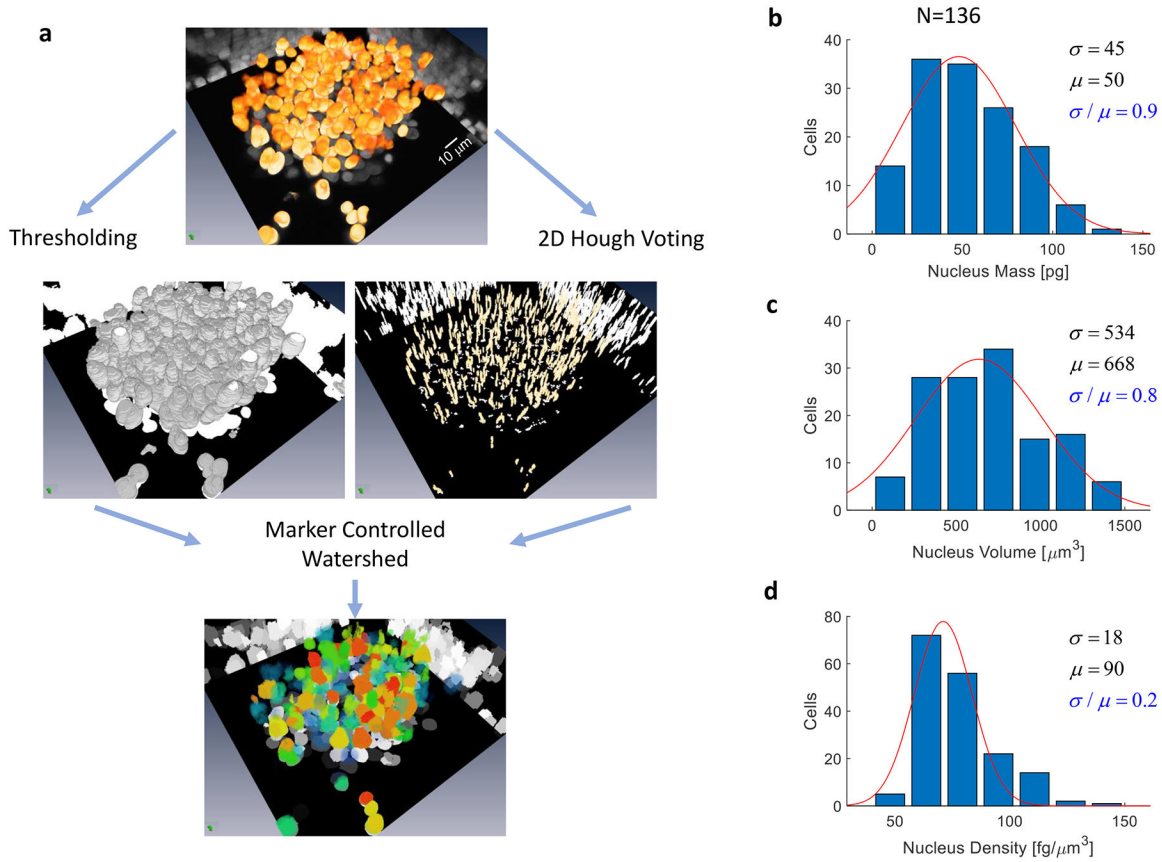


Figure 5.

Automated segmentation of cells inside spheroids. **a**, Instance segmentation of spheroids was performed by 3D marker-controlled watershed on the ACM estimated DNA signal, with markers being determined through 2D Hough voting on a per-z-slice basis. The result of the Hough voting is a volume with a unique marker on the spheroid, which resembles a column tracking the center of the nucleus through the focus. The result of watershed is a 3D volume with a unique label for each nucleus within the spheroid, which enables the calculation of parameters on individual cells. **b**, Distribution of nuclear dry mass. **c**, Distribution of nuclear volume. **d**, Distribution of dry mass density. The standard deviation (σ), mean (μ), and their ratios are indicated for each plot. Note that the nuclear mass density (**d)** has a significantly narrower distribution compared to nuclear mass (**b)**, and volume (**c)**, as indicated by the σ/μ ratio.