Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy

Chawin Ounkomol[1], Sharmishtaa Seshamani[2], Mary M. Maleckar[1], Forrest Collman[2] and Gregory R. Johnson[1]*

Understanding cells as integrated systems is central to modern biology. Although fluorescence microscopy can resolve subcellular structure in living cells, it is expensive, is slow, and can damage cells. We present a label-free method for predicting three-dimensional fluorescence directly from transmitted-light images and demonstrate that it can be used to generate multi-structure, integrated images. The method can also predict immunofluorescence (IF) from electron micrograph (EM) inputs, extending the potential applications.

Imaging methods currently used to capture details of cellular organization all present restrictions with respect to expense, spatio-temporal resolution, and sample perturbation. Fluorescence microscopy permits imaging of structures of interest by specific labeling, but can require advanced instrumentation and time-consuming sample preparation. Significant phototoxicity and photobleaching can also perturb samples, creating a tradeoff between data quality and time scales available for live cell imaging[1,2]. Furthermore, the number of simultaneous fluorescent tags is restricted by both spectrum saturation and cell health, limiting the number of parallel labels for joint imaging. In contrast, transmitted-light microscopy (TL), for example, bright-field, phase, differential interference contrast imaging is relatively low cost and is label free (greatly reduced phototoxicity[3] and simplified sample preparation). Although TL images provide valuable information about cellular organization, they lack the clear contrast of fluorescence labeling. EMs also contain rich biological detail about subcellular structure, but often require tedious expert interpretation. A method combining the clarity of fluorescence microscopy with the relative simplicity and modest cost of other imaging techniques would present a ground-breaking tool for obtaining biological insight into the integrated organization of subcellular structures.

Convolutional neural networks (CNNs) capture nonlinear relationships over large areas of images, resulting in vastly improved performance for image recognition tasks as compared to classical machine learning methods. Here we present a CNN-based tool, employing a U-Net architecture[4] (Supplementary Fig. 1, Methods) to model relationships between distinct but correlated imaging modalities, and show the efficacy of this tool for predicting corresponding fluorescence images directly from both three-dimensional (3D) TL live cell images and two-dimensional (2D) EMs alone.

The label-free prediction tool learns each relationship between 3D TL and fluorescence live cell images for several major subcellular structures (Fig. 1a–c; cell membrane, DNA, etc.). A resultant model can then predict a 3D fluorescence image from a new TL input. A single TL input can be applied to multiple subcellular structure models, enabling multi-channel, integrated fluorescence imaging (Fig. 1d,e). The method can similarly be used to predict 2D IF images directly from EMs to highlight distinct subcellular structures and to register conjugate multi-channel fluorescence data with EMs[5] (Fig. 2).

In our experiments, we used only spatially registered pairs of images from a relatively small set (30 pairs per structure for 3D TL-to-fluorescence, and 40 for 2D EM-to-fluorescence; Methods) to train each structure model. Biological detail observed in predictions varies among subcellular structures modeled; however, in the case of the 3D TL-to-fluorescence models, predictions appear structurally similar to ground-truth fluorescence. Nuclear structures are well resolved; for example, images produced by the DNA model (Fig. 1b) depict well-formed and separated nuclear regions, as well as finer detail, including chromatin condensation just before and during mitosis, and the nuclear envelope model predictions (Supplementary Fig. 2) provide high-resolution localization and 3D morphology. The nucleoli model also resolves the precise location and morphology of individual nucleoli (Supplementary Fig. 2).

TL-to-fluorescence models' performance was quantified by the Pearson correlation coefficient on 20 predicted and corresponding ground-truth fluorescence image pairs (Fig. 1c) from independent test sets for each model (Methods). A theoretical upper bound for model performance based on an estimate of signal-to-noise ratio (SNR) of fluorescence images used for training was determined (Methods). Model performance for each structure is well bounded by this limit (Fig. 1c).

We trained a model predicting DNA with an extended procedure ("DNA+"; Methods) to evaluate whether predictions improve with additional training images and iterations. Outcome improved as measured by an increase in Pearson correlation, and images qualitatively showed better clarity of subnuclear structure and precision of predictions around mitotic cells (Fig. 1c, Supplementary Fig. 2, Methods). Most critically, these details can be observed together in a 3D integrated multi-channel prediction derived from a single TL image (Fig. 1e and Supplementary Video). Examples for all 14 labeled structure models' predictions on each model's test set can be found in Supplementary Fig. 2.

Transforming one imaging modality into another also can be useful in less direct ways: 2D IF images predicted from EMs (Fig. 2) can be used to facilitate automatic registration of conjugate multi-channel fluorescence data with EMs. Array tomography data[5] of ultra-thin brain sections uses EMs and ten channels of IF images (including the structure of myelin basic protein, MBP-IF) from the

[1]Allen Institute for Cell Science, Seattle, WA, USA. [2]Allen Institute for Brain Science, Seattle, WA, USA. *e-mail: gregj@alleninstitute.org
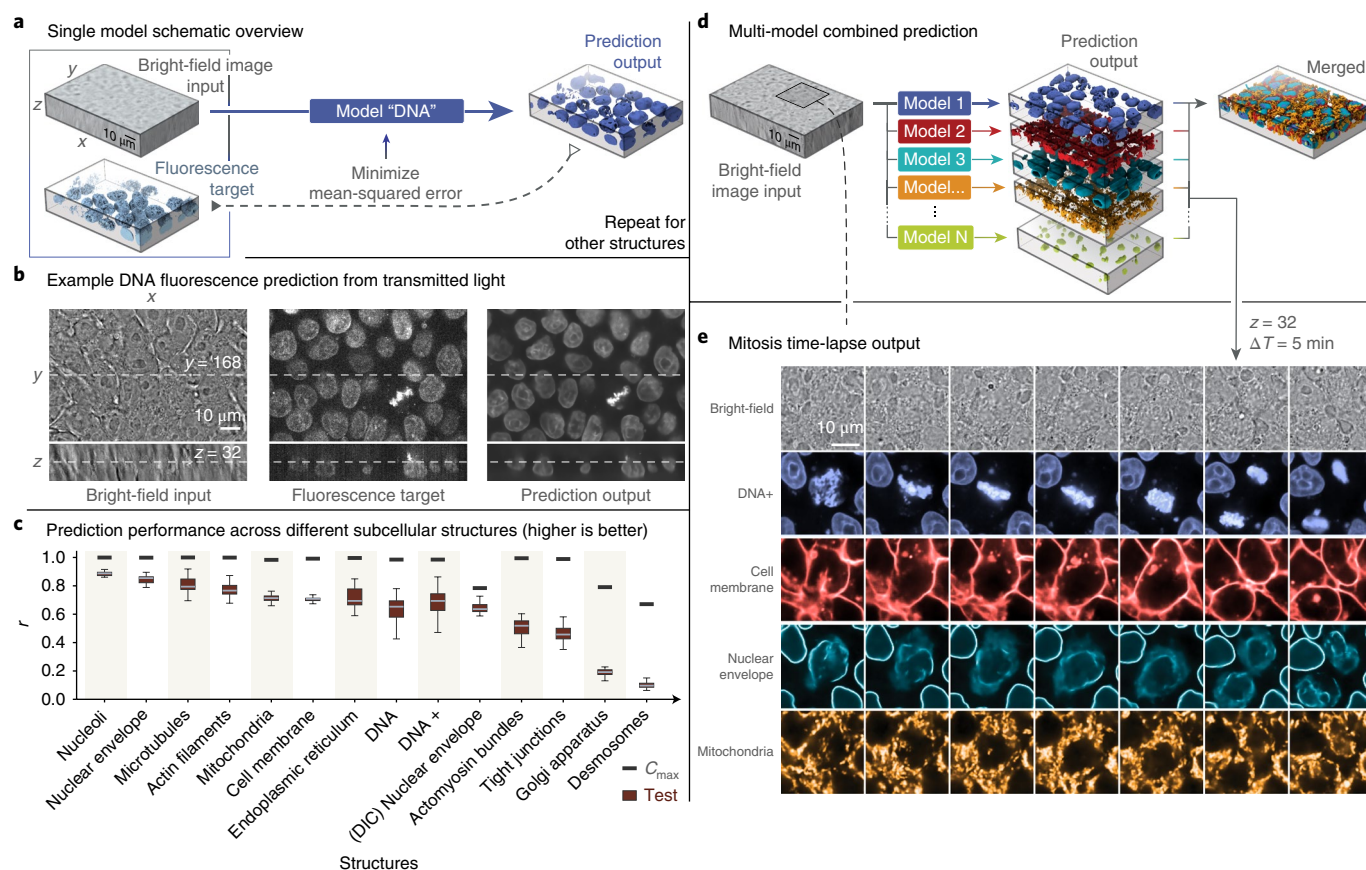
Fig. 1 | Label-free imaging tool pipeline and application using 3D transmitted light-to-fluorescence models. a, Given transmitted light and fluorescence image pairs as input, the model is trained to minimize the mean-squared error (MSE) between the fluorescence ground truth and output of the model. **b**, Left to right: an example of a 3D input transmitted-light image, a ground-truth confocal DNA fluorescence image, and a tool prediction. **c**, Distributions of the image-wise Pearson correlation coefficient (*r*) between ground-truth (target) and predicted test images derived from the indicated subcellular structure models. Each target/predicted image pair in the test set is a point in the resultant *r* distribution; the 25th, 50th, and 75th percentile image pairs are spanned by the box for each indicated structure, with whiskers indicating the last data points within 1.5× the interquartile range of the lower and upper quartiles. The number of images (*n*) was 18 for the cell membrane, 10 for the differential interference contrast (DIC) nuclear envelope, and 20 for all other distributions. A complete description of the structure labels is provided in the Methods. Black bars indicate maximum correlation between the target image and a theoretical, noise-free image (*C*$_{max}$; details in the Methods). **d**, Individual subcellular structure models are applied to the same input and combined to predict multiple structures. **e**, Localization of DNA (blue), cell membrane (red), nuclear envelope (cyan), and mitochondria (orange) as predicted for time-lapse transmitted-light (bright-field) input images taken at 5-min intervals (center *z*-slice shown); a mitotic event with stereotypical reorganization of subcellular structures is clearly evident. Similar results were observed for two independent time-series input image sets. All results shown here are obtained from new transmitted-light images not used during model training.

same sample, but from two different microscopes. Thus, EM and IF images are not natively spatially aligned. Although EMs and corresponding images from other modalities can be registered by hand, resulting in multi-channel conjugate EM images[5–7], manual registration is tedious and time-consuming. We trained a 2D version of the label-free tool on manually registered pairs of EM and MBP-IF images and then used model predictions to register an EM image ($15 \times 15 \, \mu m^2$) to a much larger target MBP-IF image ($204.8 \times 204.8 \, \mu m^2$) (Fig. 2a). Test EM images were first input to the model to predict corresponding MBP-IF images (Fig. 2a), and conventional intensity-based matching techniques (Methods) were then used to register each MBP-IF prediction (and EM image) to the target MBP-IF image (Fig. 2b; successful convergence on 86 of 90 image pairs). The average distance between automated and ground truth registration was measured as $1.16 \pm 0.79 \, px$ (MBP-IF pixel data units). To our knowledge, this is the first successful attempt to automate this registration process via a learning-based technique, which suggests that the label-free tool's utility can be

extended to diverse imaging modalities and a variety of downstream image-processing challenges.

We next determined that individual structure models, trained solely on static images, can be used to predict fluorescence time lapse by applying several subcellular structure TL-to-fluorescence models to a single TL 3D time series (covering 95 min at 5-min intervals; Fig. 1e, Supplementary Video 1). In addition to simultaneous structure visualization, characteristic dynamics of mitotic events, including reorganization of the nuclear envelope and cell membrane, are evident in the predicted multi-channel time series (Fig. 1e). Time series acquired with a similar 5-min acquisition interval but with three-channel spinning disk fluorescence reveal both obvious bleaching artifacts and changes in cellular morphology and health after 10–15 min (data not shown). The phototoxicity occurring in extended, multi-label live-cell time-series fluorescence imaging on the human induced pluripotent stem cells (hiPSCs) used here evidences challenges in obtaining integrated structural information from fluorescence time-lapse imaging. Although many
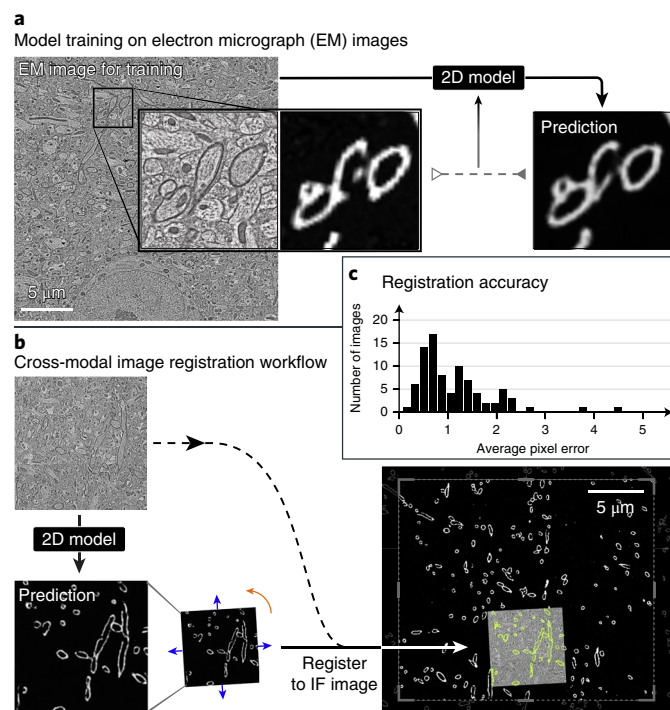
Fig. 2 | Label-free imaging tool facilitates 2D automated registration across imaging modalities. We first train a model to predict a 2D myelin basic protein immunofluorescence image (MBP-IF) from a 2D EM and then register this prediction to automate cross-modal registration.
a, An example EM with a highlighted subregion (left), the MBP-IF image corresponding to the same subregion (middle), and the label-free imaging tool prediction of the same subregion given only the EM as input (right).
b, The EM of the subregion to be registered (top left) is passed through the trained 2D model to obtain a prediction for the subregion (bottom left), which is then registered to MBP-IF images within a larger field of view (bottom right) (see Methods for details). Only a $20 \times 20\,\mu m^2$ region from the $204.8 \times 204.8\,\mu m^2$ MBP-IF search image is shown; predicted and registered MBP-IF are overlaid (in green) together with the EM image.
c, Histogram of average distance between automated registration and manual registration as measured across 90 test images, in units of pixels of MBP-IF data. This distribution has an average of $1.16 \pm 0.79\,px$, where manual registrations between two independent annotators differed by $0.35 \pm 0.2\,px$.

strategies exist to minimize this photodamage (that is, oxygen scavenging, reduced laser power and exposure, advanced microscopy techniques[2], and machine-learning-driven denoising[8]), all require compromises with respect to ease, image quality, and fidelity. This method avoids trade-offs and directly produces time-series predictions for which no fluorescence imaging ground truth exists, thereby greatly increasing the time scales over which some cellular processes can be visualized and measured.

Our method has inherent limitations. Models must learn a relationship between distinct but correlated imaging modes; predictive performance is contingent on the existence of this association. In the case of desmosomes or actomyosin bundles, for example, model performance for the presented training protocol was poor, presumably owing to a weaker association between TL and fluorescence images of these structures (Fig. 1c, Supplementary Fig. 2). The quality and quantity of training data also influence the accuracy of model predictions, although this relationship is highly nonlinear in tested cases (for DNA model performance, we saw diminishing returns between 30 and 60 images; Supplementary Fig. 3).

Performance between models varied with 2D versus 3D information: use of a 2D DNA model to predict z-slices selected from 3D images showed artifacts between predicted z-slices and decreased correlation between ground truth and predictions (Supplementary Fig. 4; Methods), suggesting that 3D interference patterns are valuable for predicting subcellular organization.

Additionally, we cannot assess a priori how models will perform in contexts for which there are very few or no examples in training or testing data. Models pre-trained using one cell type (that is, hiPSC) did not perform as well with inputs of drastically different cellular morphologies (Supplementary Fig. 5). We compared predictions from the DNA+ model (trained on hiPSC images) to those from a model trained on images of DNA-labeled human embryonic kidney (HEK) 293 kidney-phenotype cells[9] (applied to both hiPSC and HEK293 test images). Although gross image features were comparable, prediction performance for morphological detail improved markedly when the model was trained on and applied to data of the same cell type. A similar reduction in predictive performance was evident when pre-trained models were used to predict a fluorescent DNA label in different cell types, like cardiomyocytes or HT-1080 fibroblast-phenotype cells[10] (Supplementary Fig. 5, Methods).

Furthermore, predictions from inputs acquired with imaging parameters identical to those used to compose models' training sets will provide the most accurate results versus ground-truth data. For example, we observed decreased model accuracy when predicting fluorescence images from input TL stacks acquired with a shorter inter-slice interval (~0.13 s) than that in training data (~2.2 s) (data not shown). Ultimately, when evaluating the utility of predicted images, one must consider the context of use. For instance, DNA or nuclear membrane predictions may have sufficient accuracy for application to downstream nuclear segmentation algorithms, but microtubule predictions would not be effective for assaying rates of microtubule polymerization (Fig. 1e, Supplementary Fig. 2). Finally, there may not be a direct quantitative link between the predicted intensity of a tagged structure and protein levels.

The presented methodology has wide potential use in many biological imaging fields. Primarily, it may reduce or even eliminate routine capture of some images in existing imaging and analysis pipelines, permitting similar throughput in a more efficient, cost-effective manner. Notably, training data requires no manual annotation, little to no pre-processing, and relatively small numbers of paired examples, drastically reducing the barrier to entry associated with some machine learning approaches. This approach may have particular value in image-based screens where cellular phenotypes can be detected via expressed fluorescent labels[11], pathology workflows requiring specialized labels that identify specific tissues[12], and long time-series observation of single cells[1], tissues, or organism-level populations where more expensive instrumentation is not available[2]. Recent related work convincingly demonstrates that 2D whole-cell antibody stains can be predicted from TL[13], supporting the conclusion that similar techniques can be applied to a wide variety of biological studies, as demonstrated here by automatic registration of conjugate multi-channel fluorescence data with EMs. The method is additionally promising when generating a complete set of simultaneous ground-truth labels is infeasible, for example, in live-cell time-series imaging. Finally, our tool permits the generation of integrated images by which multi-dimensional interactions among cellular components can be investigated. This implies exciting potential for probing coordination of subcellular organization as cells grow, divide, and differentiate, and signifies a new opportunity for understanding structural phenotypes in the context of disease modeling and regenerative medicine. More broadly, the presented work suggests an opportunity for a key new direction in biological imaging research: the exploitation of imaging modalities' indirect but learnable relationships to visualize biological features of interest with ease, low cost, and high fidelity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41592-018-0111-2.

References

1. Skylaki, S., Hilsenbeck, O. & Schroeder, T. Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol.* **34**, 1137–1144 (2016).
2. Chen, B.-C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
3. Selinummi, J. et al. Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images. *PLoS One* **4**, e7497 (2009).
4. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-assisted Intervention—MICCAI 2015* (eds Navab, N. et al.) 234–241 (Springer, Cham, 2015).
5. Collman, F. et al. Mapping synapses by conjugate light-electron array tomography. *J. Neurosci.* **35**, 5792–5807 (2015).
6. Russell, M. R. G. et al. 3D correlative light and electron microscopy of cultured cells using serial blockface scanning electron microscopy. *J. Cell. Sci.* **130**, 278–291 (2017).
7. Kopek, B. G., Shtengel, G., Grimm, J. B., Clayton, D. A. & Hess, H. F. Correlative photoactivated localization and scanning electron microscopy. *PLoS One* **8**, e77209 (2013).
8. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/07/03/236463 (2018).
9. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**, 59–74 (1977).
10. Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. & Gardner, M. B. Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* **33**, 1027–1033 (1974).
11. Goshima, G. et al. Genes required for mitotic spindle assembly in *Drosophila* S2 cells. *Science* **316**, 417–421 (2007).
12. Gurcan, M. N. et al. Histopathological image analysis: a review. *IEEE. Rev. Biomed. Eng.* **2**, 147–171 (2009).
13. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 (2018).

Author contributions

G.R.J. conceived the project. C.O. implemented the model for 2D and 3D images. M.M. provided guidance and support. C.O., S.S., F.C., and G.R.J. designed computational experiments. C.O., S.S., M.M., F.C., and G.R.J. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0111-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to G.R.J.

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

Methods

Three-dimensional live cell microscopy. The 3D light microscopy data used to train and test the presented models consists of z-stacks of colonies of human embryonic kidney cells (HEK293)[9], human fibrosarcoma cells (HT-1080)[10], genome-edited (hiPSC) lines[10] expressing a protein endogenously tagged with either monomeric enhanced green fluorescent protein (mEGFP) or red fluorescent protein (mTagRFP) that localizes to a particular subcellular structure[14], and hiPSC-derived cardiomyocytes differentiated from the former. The EGFP-tagged proteins and their corresponding structures are: α-tubulin (microtubules), β-actin (actin filaments), desmoplakin (desmosomes), lamin B1 (nuclear envelope), fibrillarin (nucleoli), myosin IIB (actomyosin bundles), sec61B (endoplasmic reticulum), STGAL1 (Golgi apparatus), Tom20 (mitochondria), and ZO1 (tight junctions). The cell membrane was labeled by expression of mTagRFP tagged with a CAAX motif.

Imaging. All cell types were imaged for up to 2.5 h on a Zeiss spinning disk microscope with ZEN Blue 2.3 software and with a 100×/1.25-NA (numerical aperture) objective (Zeiss C-Apochromat ×100/1.25 W Corr), with up to four 16-bit data channels per image: transmitted light (either bright-field or DIC), cell membrane labeled with CellMask, DNA labeled with Hoechst, and EGFP-tagged cellular structure. Respectively, acquisition settings for each channel were as follows: white LED, 50-ms exposure; 638-nm laser at 2.4 mW, 200-ms exposure; 405 nm at 0.28 mW, 250-ms exposure; 488-nm laser at 2.3 mW, 200-ms exposure. The exception was CAAX-RFP-based cell membrane images, which were acquired with a 63×/1.2-NA objective (Zeiss C-Apochromat ×63/1.2 W Corr), a 561-nm laser at 2.4 mW, and a 200-ms exposure. Z-slice images taken with a ×100× objective were captured at a YX resolution of 624×924 px^2 with a pixel scale of 0.108 μm px^{-1}, and 63×-objective z-slice images were captured at a YX resolution of $1,248 \times 1,848$ px^2 with a pixel scale of 0.086 μm px^{-1}. All z-stacks were composed of 50–75 z-slices with an inter-z-slice interval of 0.29 μm. Images of cardiomyocytes contained 1–5 cells per image, whereas images of other cell types contained 10–30 cells per image. Time-series data were acquired using the same imaging protocol as for acquisition of training data but on unlabeled, wild-type hiPSCs at 5-min intervals for 95 min, with all laser powers set to zero to reproduce the inter-z-slice timing of the training images.

Tissue culture. hiPSCs, HEK293 cells, or HT-1080 cells were seeded onto Matrigel-coated 96-well plates at the densities specified below. The cells were stained on the days they were to be imaged, first by incubation in their imaging media with ×1 NucBlue (Hoechst 33342, Thermo Fisher) for 20 min. hiPSCs were then incubated in imaging media with ×1 NucBlue and ×3 CellMask (Thermo Fisher) for 10 min, whereas HEK293 and HT-1080 cells were then incubated in imaging media with ×1 NucBlue and ×0.5 CellMask for 5 min. The cells were washed with fresh imaging media before imaging.

For hiPSCs, the culture media was mTeSR1 (Stem Cell Technologies) with 1% penicillin–streptomycin (Pen-Strep). The imaging media was phenol-red-free mTeSR1 with 1% Pen-Strep. Cells were seeded at a density of ~2,500 cells per well and were imaged 4 d after initial plating. For HEK293 cells, the culture media was Dulbecco's modified Eagle's medium (DMEM) with GlutaMAX (Thermo Fisher), 4.5 g l^{-1} D-glucose, 10% fetal bovine serum (FBS), and 1% antibiotic-antimycotic. The imaging media was phenol-red-free DMEM/F-12 with 10% FBS and 1% antibiotic-antimycotic. Cells were seeded at a density of 13,000–40,000 cells per well and were imaged 1–2 d after initial plating. For HT-1080 cells, the culture media was DMEM with GlutaMAX, 15% FBS, and 1% Pen-Strep. The imaging media was phenol-red-free DMEM/F-12 with 10% FBS and 1% Pen-Strep. Cells were seeded at a density of 2,500–40,000 cells per well and were imaged 4 d after initial plating.

CAAX-tagged hiPSCs were differentiated to cardiomyocyte phenotype by seeding onto Matrigel-coated six-well tissue culture plates at a density ranging from 0.15 to 0.25×10^6 cells per well in mTeSR1 supplemented with 1% Pen-Strep, 10 μM ROCK inhibitor (Stem Cell Technologies) (day −3). Cells were grown for 3 d with daily mTeSR1 changes. On day 0, we initiated differentiation by treating cultures with 7.5 μM CHIR99021 (Cayman Chemical) in Roswell Park Memorial Institute (RPMI) media (Invitrogen) containing insulin-free B27 supplement (Invitrogen). After 2 d, cultures were treated with 7.5 μM IWP2 (R&D Systems) in RPMI media with insulin-free B27 supplement. On day 4, cultures were treated with RPMI with insulin-free B27 supplement. From day 6 onward, media was replaced with RPMI media supplemented with B27 containing insulin (Invitrogen) every 2–3 d. Cardiomyocytes were re-plated at day 12 onto glass-bottom plates coated with PEI/laminin and were imaged on day 43 after initiation of differentiation. The imaging media was phenol-red-free RPMI with B27. Prior to imaging, cells were stained by incubation in imaging media with Nuclear Violet (AAT Bioquest) at a 1/7,500 dilution and ×1 CellMask for 1 min and then washed with fresh imaging media.

Data for training and evaluation. Supplementary Table 1 outlines the data used to train and evaluate the models based on 3D live-cell z-stacks, including train-test data splits. All multi-channel z-stacks were obtained from a database of images produced by the Allen Institute for Cell Science's microscopy pipeline (see http://www.allencell.org). For each of the 11 hiPSC cell lines, we randomly selected z-stacks from the database and paired the transmitted-light channel with the EGFP/

RFP channel to train and evaluate models (Fig. 1c) to predict the localization of the tagged subcellular structure. The transmitted-light channel modality was bright-field for all but the DIC-to-nuclear envelope model. For the DNA model data, we randomly selected 50 z-stacks from the combined pool all bright-field-based z-stacks and paired the transmitted-light channel with the Hoechst channel. The training set for the DNA+ model was further expanded to 540 z-stacks with additional images from the Allen Institute for Cell Science's database. Note that while a CellMask channel was available for all z-stacks, we did not use this channel because the CAAX-membrane cell line provided higher quality images for training cell membrane models. A single z-stack time series of wild-type hiPSCs was used only for evaluation (Fig. 1e).

For experiments testing the effects of the number of training images on model performance (Supplementary Fig. 3), we supplemented each model's training set with additional z-stacks from the database. Z-stacks of HEK293 cells were used to train and evaluate DNA models, whereas all z-stacks of cardiomyocytes and of HT-1080 cells were used only for evaluation (Supplementary Fig. 5). The 2D DNA model (Supplementary Fig. 4) used the same data as the DNA+ model.

All z-stacks were converted to floating-point and were resized via cubic interpolation such that each voxel corresponded to $0.29 \times 0.29 \times 0.29$ μm^3, and resulting images were 244×366 px^2 for 100×-objective images or 304×496 px^2 for 63×-objective images in Y and X, respectively, and between 50 and 75 pixels in Z. Pixel intensities of all input and target images were z-scored on a per-image basis to normalize any systematic differences in illumination intensity.

Electron and immunofluorescence microscopy. *Imaging.* For conjugate array tomography data[5], images of 50 ultra-thin sections were taken with a wide-field fluorescence microscope using three rounds of staining and imaging to obtain ten-channel IF data (including MBP) at 100 nm per pixel. Five small regions were then imaged with a field emission scanning electron microscope to obtain high-resolution EMs at 3 nm per pixel. Image processing steps independently stitched the IF sections and one of the EM regions to create 2D montages in each modality. Each EM montage was then manually registered to the corresponding MBP channel montage with TrakEM2[15].

Data used for training and evaluation. Forty pairs of registered EM and MBP montages were resampled to 10 nm per pixel. For each montage pair, a central region of size $3,280 \times 3,214$ px^2 was cut out and used for the resultant final training set. This corresponded to the central region of the montage that contained no unimaged regions across the sections used. Pixel intensities of the images were z-scored. For the registration task, a total of 1,500 EM images (without montaging) were used as an input to directly register to the corresponding larger MBP image in which they lay. For this, each EM image was first downsampled to 10 nm per pixel without any transformations to generate a $1,500 \times 1,500$ px^2 image.

Model architecture description and training. We employed a CNN based on the U-Net architecture[4] (Supplementary Fig. 1) because of its demonstrated performance in image segmentation and tracking tasks. In general, CNNs are uniquely powerful for image-related tasks (classification, segmentation, image-to-image regression) because they are image-translation invariant, learn complex nonlinear relationships across multiple spatial areas, circumvent the need to engineer data-specific feature extraction pipelines, and are straightforward to implement and train. CNNs have been shown to outperform other state-of-the-art models in basic image recognition[16] and have been used in biomedical imaging for a wide range of tasks including image classification, object segmentation[17], and estimation of image transformations[18]. Our U-Net variant consists of layers that perform one of three convolution types, followed by a batch normalization and rectified linear unit (ReLU) operation. The convolutions are either 3-pixel convolutions with a stride of 1 pixel on zero-padded input (such that the input and output of that layer are the same spatial area), 2-pixel convolutions with a stride of 2 pixels (to halve the spatial area of the output), or 2-pixel transposed convolutions with a stride of 2 pixels (to double the spatial area of the output). There are no normalization or ReLU operations on the last layer of the network. The number of output channels per layer is shown in Supplementary Fig. 1. The 2D and 3D models use 2D or 3D convolutions, respectively.

Owing to memory constraints associated with graphics processing unit computing, we trained the model on batches of either 3D patches ($64 \times 64 \times 32$ px^3, YXZ) for light microscopy data or 2D patches (256×256 px^2) for conjugate array tomography data, which were randomly subsampled uniformly both across all training images and spatially within an image. The training procedure took place in a typical forward–backward fashion, with model parameters updated via stochastic gradient descent (backpropagation) to minimize the MSE between output and target images. All models presented here were trained using the Adam optimizer[19] with a learning rate of 0.001 and with beta values of 0.5 and 0.999 for 50,000 mini-batch iterations. We used a batch size of 24 for 3D models and of 32 for 2D models. Running on a Pascal Titan X, each model completed training in approximately 16 h for 3D models (205 h for DNA+) and in 7 h for 2D models. Training of the DNA+ model was extended to 616,880 mini-batch iterations. For prediction tasks, we minimally crop the input image such that its size in any dimension is a multiple of 16, to accommodate the multi-scale aspect of the CNN architecture. Prediction

takes approximately 1 s for 3D images and 0.5 s for 2D images. Our model training pipeline was implemented in Python using the PyTorch package (http://pytorch.org).

Three-dimensional light microscopy model results analysis and validation. For 3D light microscopy applications, model accuracy was quantified by the Pearson correlation coefficient,

$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 \sum (y - \overline{y})^2}}$$

between the pixel intensities of the model's output, y, and independent ground-truth test images, x (Fig. 1c, Supplemental Figs. 3, 4b, 5b, and 6). To estimate the theoretical upper bound on the performance of a model, we calculated the correlation for a theoretical model that is able to perfectly predict the spatial fluctuations of the signal but is unable to predict the random fluctuations in the target image that arise from fundamentally unpredictable phenomena (such as noise in the electronics of the camera or fluctuations in the number of photons collected from a fluorescent molecule). Intuitively, as the size of random fluctuations increases relative to the size of the predictable signal, one would expect the performance of even a perfect model to degrade. The images in Fig. 1 of DNA-labeled targets and predictions make this point, in so far as the model cannot be expected to predict the background noise in the DNA-labeled imagery. Therefore, to estimate a lower bound on the amplitude of the random fluctuations, we analyzed images of cells that were taken with identical imaging conditions but contained no fluorescent labels, for example, images taken with microscope settings designed to detect Hoechst staining, but with cells for which there was no Hoechst dye applied, or images taken with microscope settings designed to detect GFP but with cells with no GFP present. We used the variance of pixel intensities across the image as an estimate of the variance of random fluctuations (N), and then averaged that variance across control images in order to arrive at our final estimate. Calculating the correlation between a perfect model prediction S (equal to the predictable image) and an image T, which is the combination of the predictable image and the random fluctuations ($T_{x,y,z} = N_{x,y,z} + S_{x,y,z}$), is

$$C_{\max} = \sqrt{\frac{\mathrm{SNR}}{1 + \mathrm{SNR}}} \quad \text{where } \mathrm{SNR} = \frac{\langle S^2 \rangle}{\langle N^2 \rangle}$$

If we assume the correlation between the predictable component and the random component is zero, then the variance of the predictable image ($\langle S^2 \rangle$) can be calculated by taking the variance of the measured image ($\langle T^2 \rangle$) and subtracting the variance of the random fluctuations ($\langle N^2 \rangle$). The result is a formula for the theoretical upper bound of model performance that depends only on the lower-bound estimate of the variance of the noise and the variance of the measured image. We report the average value of $C_{\max}$ for all images in the collection as black tick marks in Fig. 1c.

Registration across imaging modalities. We employed a 2D version of our tool trained on the montage pairs described below. EM images were reflection padded to $1{,}504 \times 1{,}504 \text{ px}^2$ and passed through the trained model, and then predictions were cropped back to the original input size to generate an MBP prediction image. This MBP prediction image was first roughly registered to the larger MBP-IF images using cross-correlation-based template matching for a rigid transformation estimate. Next, the residual optical flow[20] between the predicted image transformed by the rigid estimate and the MBP-IF image was calculated, which was then used to fit a similarity transformation that registers the two images, implemented using OpenCV (www.opencv.org). Ninety prediction images were randomly selected from the larger set, where more than 1% of the predicted image pixels were greater than 50% of the maximum intensity, to ensure that the images contained sufficient MBP content to drive registration. Ground-truth transformation parameters were calculated by two independent authors on this subset of EM images by manual registration (3–4 min per pair) to the MBP-IF images using TrakEM2. Because the images were registered using a similarity transformation where it is possible for the registration accuracy of the central pixels and those at the edges to be different, we calculated the registration errors by computing the average difference in displacement across an image, as measured in pixels of the target IF image. We report these results for registration differences (Fig. 2) between authors and between the algorithm estimate and one of the authors.

Three-dimensional fluorescence image predictions from a 2D model. To compare performance between models trained on 2D and 3D data, we trained a 2D DNA model for evaluation against the DNA+ model. The 2D model was trained on the same dataset with the same training parameters as the DNA+ with the exception that training patches of size $64 \times 64 \text{ px}^2$ were sampled from random z-slices of the 3D training images. The model was trained for 250,000 mini-batch iterations with a batch size of 24 for a total training time of approximately 18 h. After training, we formed 3D predicted fluorescence images by inputting sequential 2D bright-field z-slices into the model and combining the outputs into 3D volumes (Supplementary Fig. 4).

Reporting Summary. Further information on research design can be found in the Nature Research Reporting Summary linked to this article.

Code availability. Software for training and using trained models is available at https://github.com/AllenCellModeling/pytorch_fnet/tree/release_1.

Data availability

Data used to train the 3D models are available at https://downloads.allencell.org/publication-data/label-free-prediction/index.html.

References

14. Roberts, B. et al. Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* **28**, 2854–2874 (2017).
15. Cardona, A. et al. TrakEM2 software for neural circuit reconstruction. *PLoS One* **7**, e38011 (2012).
16. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
17. Zhou, S. K., Greenspan, H. & Shen, D. *Deep Learning for Medical Image Analysis* (Academic Press, Cambridge, MA, 2017).
18. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image. Anal.* **42**, 60–88 (2017).
19. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv* Preprint at https://arxiv.org/abs/1412.6980 (2014).
20. Farnebäck, G. Two-frame motion estimation based on polynomial expansion. In *Image Analysis: SCIA 2003* (eds Bigun, J. & Gustavsson, T.) 363–370 (Springer, Berlin, Heidelberg, 2003).

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

natureresearch

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	3D confocal images were acquired via Zeiss ZEN Blue 2.3 software, Electron micrographs were acquired with Zeiss Atlas - Large Area Imaging
Data analysis	Custom to reproduce the results presented in the paper can be found here: https://github.com/AllenCellModeling/pytorch_fnet/tree/release_1, Pytorch 0.4, Open CV, TrakEM2 1.0i

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that was used to train the 3D models is available at http://downloads.allencell.org/publication-data/label-free-prediction/index.html

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by picking a low number of training image pairs (once), and observing that our model performed well by chosen metrics.
Data exclusions	Images were randomly selected from a corpus of production images for a large survey of subcellular localization in induced human pluripotent stem cells as a corpus of images for this study. This was done to acquire a minimally biased set of data.
Replication	Software and data to reproduce results is available for download. Instructions are provided in the software repository for training models from scratch. We have re-trained models several times and were able to reproduce our results.
Randomization	Samples were randomly allocated to a training set.
Blinding	Group allocation into train and test sets, was performed deterministically by a seeded random number generator, and test data was not evaluated upon completion of the training procedure.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All cell lines used for live imaging currently are, or will be available at http://www.allencell.org/cell-catalog.html

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Cell lines used were AICS-15 (Nucleoli), AICS-13 (Nucler envelope), AICS-7 (Microtubules), AICS-7 (Actin filaments), AICS-11 (Mitochondria), AICS-54 (Cell membrane), AICS-10 (Endoplasmic reticulum), AICS-24 (Actomyosin bundles), AICS-23 (Tight junctions), AICS-25 (Golgi apparatus), AICS-17 (Desomosomes), AICS-0 (Wildtype). The AICS source WTC line was provided by Bruce R. Conklin, at The Gladstone Institutes. HEK-293 and HT-1080 cells were purchased from ATCC.
Authentication	We did not authenticate the HEK-293 and HT-1080 lines. Edited AICS Cell Lines were authenticated as follows: Genetic screening with ddPCR During clone expansion, a sample of cells is pelleted for total gDNA extraction using the PureLink Pro 96 Genomic DNA Purification Kit (Life Technologies). ddPCR is performed using the Bio-Rad QX200 Droplet Reader, Droplet Generator, and QuantaSoft software. The reference assay for the 2-copy, autosomal gene RPP30 is purchased from Bio-Rad. The assay for

mEGFP detection is as follows: primers (5'-GCCGACAAGCAGAAGAACG-3', 5'-GGGTGTTCTGCTGGTAGTGG-3') probe (/56-FAM/AGATCCGCC/ZEN/ACAACATCGAGG/3IABkFQ/). The assay for AMP is as follows: primers (5'- TTTCCGTGTCGCCCTTATTCC -3', 5'- ATGTAACCCACTCGTGCACCC -3') probe (/5HEX/TGGGTGAGC/ZEN/AAAAACAGGAAGGC/3IABkFQ/). The reported final copy number of mEGFP per genome is calculated as the ratio of [(copies/μLmEGFP)-(copies/μLnonintegraedAMP)]/(copies/μLRPP30), where a ratio of 0.5 ~ 1 copy per genome and a ratio of 1 ~ 2 copies/genome. The AMP sequence is only used to normalize mEGFP signal when integration into the genome is ruled out during primary screening. For primary screening [(copies/μLmEGFP)/(copies/μLRPP30)] is plotted against [(copies/μLAMP)/(copies/μLRPP30)] in order to identify cohorts of clones for ongoing analysis.

Genetic screening with tiled junctional PCR
PCR was used to amplify the tagged allele in two tiled reactions spanning the left and right homology arms, the mEGFP and linker sequence, and portions of the distal genomic region 5' of the left homology arm and 3' of the right homology arm using PrimeStar® (Clontech) PCR reagents and gene-specific primers. Both tiled junctional PCR products were Sanger sequenced bi-directionally with PCR primers when their size was validated by gel electrophoresis and/or fragment analysis (Advanced Analytics Fragment Analyzer).

Screening for clones with wild type untagged allele sequences
PCR is also used to amplify the untagged allele using gene-specific primers. These primers do not selectively amplify the unmodified locus, as is the case for tiled junctional PCR amplification of the tagged allele, but rather amplified both untagged and tagged alleles. Tracking of insertions and deletions (INDELs) by decomposition (TIDE) analysis is performed manually on the amplification reaction after bidirectional Sanger sequencing in order to determine the sequence of the untagged allele. For all final clones with wild type untagged alleles, the PCR product corresponding to the untagged allele is gel isolated and sequenced to confirm the initial result from TIDE analysis.

Off-target PCR screening
Cas-OFFinder is used to identify potential off-targets (NRG PAMs with up to 3 mismatches and 1 DNA or RNA bulge) in GRCh38 genome build. Cas-OFFinder output is further filtered to identify the most problematic off-targets with the fewest number of flaws (flaw = mismatch or bulge). Problematic off-targets are defined as off-targets with up to one flaw in the seed region (10 nts at 3' end) and up to 2 flaws in the non-seed region (10 nts at 5' end) with an NGG or NAG PAM. 8-10 of these off-targets are selected for sequencing with the goal of checking ~4 off-targets that fell close to exons (within 50bp) or within exons (exon feature in GRCh38 NCBI annotation 107) and ~4 off-targets that are closest in sequence to on-target crRNA. Approximately 300bp of sequence flanking each off-target is amplified by PCR and Sanger sequenced.

Mycoplasma contamination	All cell lines were tested negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	HEK lines have been known to be contaminated with HeLa