scientific reports

OPEN

Machine learning-based estimation of spatial gene expression pattern during ESC-derived retinal organoid development

Yuki Fujimura[1], Itsuki Sakai[2], Itsuki Shioka[2], Nozomu Takata[3], Atsushi Hashimoto[4], Takuya Funatomi[1]✉ & Satoru Okuda[2,5]✉

Organoids, which can reproduce the complex tissue structures found in embryos, are revolutionizing basic research and regenerative medicine. In order to use organoids for research and medicine, it is necessary to assess the composition and arrangement of cell types within the organoid, i.e., spatial gene expression. However, current methods are invasive and require gene editing and immunostaining. In this study, we developed a non-invasive estimation method of spatial gene expression patterns using machine learning. A deep learning model with an encoder-decoder architecture was trained on paired datasets of phase-contrast and fluorescence images, and was applied to a retinal organoid derived from mouse embryonic stem cells, focusing on the master gene Rax (also called Rx), crucial for eye field development. This method successfully estimated spatially plausible fluorescent patterns with appropriate intensities, enabling the non-invasive, quantitative estimation of spatial gene expression patterns within each tissue. Thus, this method could lead to new avenues for evaluating spatial gene expression patterns across a wide range of biology and medicine fields.

Stem cell-based tissue engineering has become essential for understanding organogenesis and has gained more attention in regenerative medicine[1–7]. For example, the loss of our body parts due to injury, aging, or disease can lead to serious life-threatening conditions, which could be slowed down or reduced by cell transplantation therapy[8,9]. However, autologous cell therapy using induced pluripotent stem (iPS) cells, established from patients, requires purification of cell lines and generation of high-quality target cells to avoid unwanted cell differentiation or contamination. This requirement is even more important for the transplantation of stem-cell derived tissues and organs, which require a well-structured complex of different cell types.

Recently, in vitro models derived from pluripotent stem cells, referred to as organoid, are useful for better understanding organogenesis. We have also developed several models of neuroepithelia, such as mouse embryonic stem cell (ESC)-derived retinal organoids[10–13]. These organoids recapitulate three-dimensional complex structures and developmental processes as observed in organogenesis. Organoids typically contain more cell types than tissues obtained in conventional planar differentiation culture[14,15]. Therefore, organoid technologies have the potential to serve as a source of cells, tissues, and organs for transplantation and as a screening platform for drug discovery via patient-relevant cell sources[16–18].

Grown organoids are evaluated by the expression of a master gene, such as Rax (aka Rx) for eye field, since the level of gene expression determines cell fate, and its location and timing correspond to those of tissue differentiation. To monitor the genetic activity of cells within each organoid, fluorescence-based reporters have been used along with homologous recombination-based knock-in and genome editing tools[19] (Artegiani et al.[20] from other groups that use fluorescent-based reporter systems). These reporters allow us to reveal the location and timing of target gene expression during the formation of each organoid, in which different cell types arise in a stepwise

[1]Division of Information Science, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan. [2]Division of Nano Life Science, Graduate School of Frontier Science Initiative, Kanazawa University, Kakumamachi, Kanazawa 920-1192, Japan. [3]Center for Vascular and Developmental Biology, Feinberg Cardiovascular and Renal Research Institute, Northwestern University, Chicago, IL 60611, USA. [4]OMRON SINIC X Corp., Bunkyo-ku, Tokyo 113-0033, Japan. [5]Nano Life Science Institute, Kanazawa University, Kakumamachi, Kanazawa 920-1192, Japan. ✉email: funatomi@is.naist.jp; satokuda@staff.kanazawa-u.ac.jp

manner. Moreover, since the fluorescence intensity of the reporter reflects the level of gene expression, it is also possible to quantitatively determine the spatiotemporal distribution of gene expression levels.

Although genetic analyses have revealed important genes and regulatory genomic regions, genomic alterations require the inactivation of at least one of two endogenous alleles. Therefore, a complete understanding of endogenous gene networks is missing and ethical issues preclude the use of genomic alterations for cell, tissue, and organ transplantation. Moreover, genomic alterations require substantial time and effort. The current remaining method is to manually select target regions in individual samples from brightfield or phase-contrast images, however this method has limitations of throughput and subjective classification criteria that can lead to high variability among observers. In addition to the above issues, ethical issues preclude the use of genomic alterations for cell, tissue, or organ transplantation into humans. To overcome these limitations, we developed a new method to automatically and non-invasively estimate spatial patterns of specific gene expressions in each organoid using machine learning.

Machine-learning approaches have been applied to stem cell biology, especially, isolated cells differentiated from stem cells on planar dishes[21–23]. In these studies, deep learning models were trained on a paired dataset of either brightfield or phase-contrast images and fluorescent images obtained by immunostaining. These models were able to classify individual cell types based on either brightfield or phase-contrast images. Furthermore, machine learning approaches are beginning to be applied to organoids. A pioneering work applied them to retinal organoids and succeeded in predicting whether each organoid differentiate into retina or not based on brightfield images[24]. However, estimating spatial and temporal gene expression remains challenging.

In this study, we developed a machine learning-based technique to estimate the spatiotemporal distribution of gene expression levels within each organoid, based on a phase-contrast image. We formulate the problem as an image transfer, i.e., a convolutional neural network (CNN) with an encoder-decoder architecture takes a phase-contrast image as input and directly outputs a fluorescent image with a spatial distribution of gene expression (Fig. 1A). Unlike previous methods that simply classify each cell or tissue as either differentiated or not, our method estimates pixel-wise intensities within each tissue. To demonstrate its applicability, we applied this model to a retinal organoid derived from mouse ESCs, focusing on a master gene Rax. The region expressing Rax defines an eye field and forms an optic vesicle during retinal organoid development[10,25,26]. For quantitative assessment, we used organoids derived from a cell line stably expressing the Rax reporter. Experiments demonstrated that our method succeeded in estimating spatially plausible fluorescent patterns with appropriate intensities.

Results

CNN architecture and dataset for estimating spatial gene expression patterns

Our model utilizes a CNN that takes a phase-contrast image as input and estimates a fluorescent image as output (Fig. 1A). The typical input to a CNN is a two-dimensional (2D) image. This 2D image is passed through several convolution layers, each followed by a nonlinear activation function. The training parameters correspond to the weights of these convolution kernels and the biases. Our network has a U-Net-like architecture[27], which is an encoder-decoder structure with skip connections. The embedded features from the encoder are passed through the decoder, which consists of upsampling and convolution layers to increase the resolution of the intermediate feature maps to obtain a fluorescent image as output.

In our model, the ResNet50[28] was used as the backbone of the encoder. The size of the input image for the ResNet50 is $3 \times H \times W$. To use the pre-trained model of the ResNet50, gray-scale phase-contrast images were replicated in the axis of the channel to create three-channel images. At the first layer, a convolution with stride 2 is applied to the input image to generate features of size $64 \times \frac{H}{2} \times \frac{W}{2}$. The ResNet50 has 4 residual blocks and the size of the output features of these blocks are $256 \times \frac{H}{4} \times \frac{W}{4}$, $512 \times \frac{H}{8} \times \frac{W}{8}$, $1024 \times \frac{H}{16} \times \frac{W}{16}$, and $2048 \times \frac{H}{32} \times \frac{W}{32}$, respectively. These features are then concatenated to the decoder to exploit multi-scale information. The output of the decoder is a fluorescent image of size $1 \times \frac{H}{2} \times \frac{W}{2}$. Note that each convolution layer has a batch normalization (BN) layer and a rectified linear unit (ReLU) activation function, except for the final convolution layer, which has a sigmoid activation function to constrain the range of the output values between 0 and 1.

The network was optimized by minimizing the training loss computed on the output and corresponding ground-truth fluorescent images. The combination of mean squared error (MSE) and cosine similarity, which captures structural patterns from the entire image, was used as the training loss.

To train, validate, and test our model, we cultured retinal organoids derived from mouse ESCs using the SFEBq method[10]. In this culture, a GFP gene was knocked-in under the promoter of a master gene of retinal differentiation, Rax. Using this method, we obtained a dataset of a pair of phase-contrast image and fluorescent image of Rax during retinal differentiation (Fig. 1B). Images were captured for 96 organoids at 4.5, 5, 6, 7, and 8 days after the start of SFEBq, where each sample was captured as 14 Z-stack images. This resulted in a total of $96 \times 5 \times 14 = 6720$ image pairs were obtained. These image pairs were divided into 5880, 420, and 420 samples for training, validation, and test, respectively. 84, 6, and 6 organoids were used for training, validation, and test, respectively; thus, each organoid does not appear in the different datasets. For data augmentation, we randomly flipped the input images vertically and horizontally during training. While the image resolution of both phase-contrast and fluorescent images is 960×720, the 512×512 regions where organoids appear were extracted.

Machine learning successfully estimates spatial Rax expression patterns

To demonstrate our model, we applied it to 420 samples of the test data. As a result, the proposed model successfully estimated the spatial expression patterns of Rax from phase-contrast images during retinal organoid development (Fig. 2). During development, multiple optic vesicles are formed through large and complicated deformations (Fig. 2A). This process begins with a spherical embryonic body, with some portions of the tissue surface evaginating outward to form hemispherical vesicles, i.e., optic vesicles. Importantly, the resulting

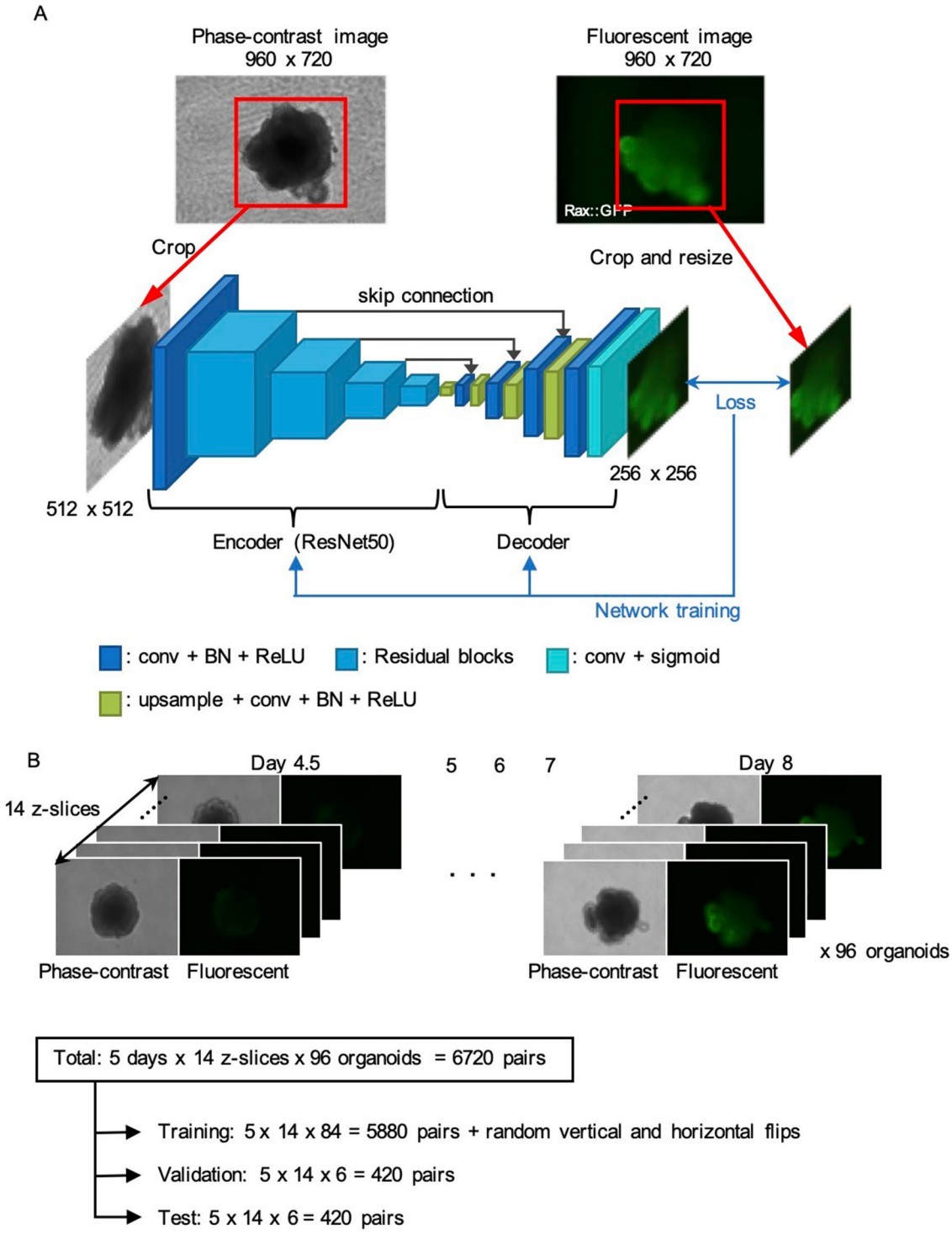

Figure 1. Network architecture and dataset for estimating spatial gene expression patterns. (**A**) Network architecture. Our network has an encoder-decoder architecture, which takes a phase-contrast image as input and then outputs a fluorescent image. The encoder backbone is the ResNet50 with four residual blocks, and the decoder has upsampling layers to increase the resolution of the intermediate feature map. (**B**) Our dataset consists of pairs of phase-contrast and fluorescent images of mouse ESC-derived retinal organoids. The fluorescence is of a GFP, which gene was knocked in under the promoter of a master gene of retinal differentiation, Rax. The dataset was divided into the training, test, and validation subsets. 84, 6, and 6 organoids were used for training, validation, and test, respectively; thus, each organoid does not appear in the different datasets.

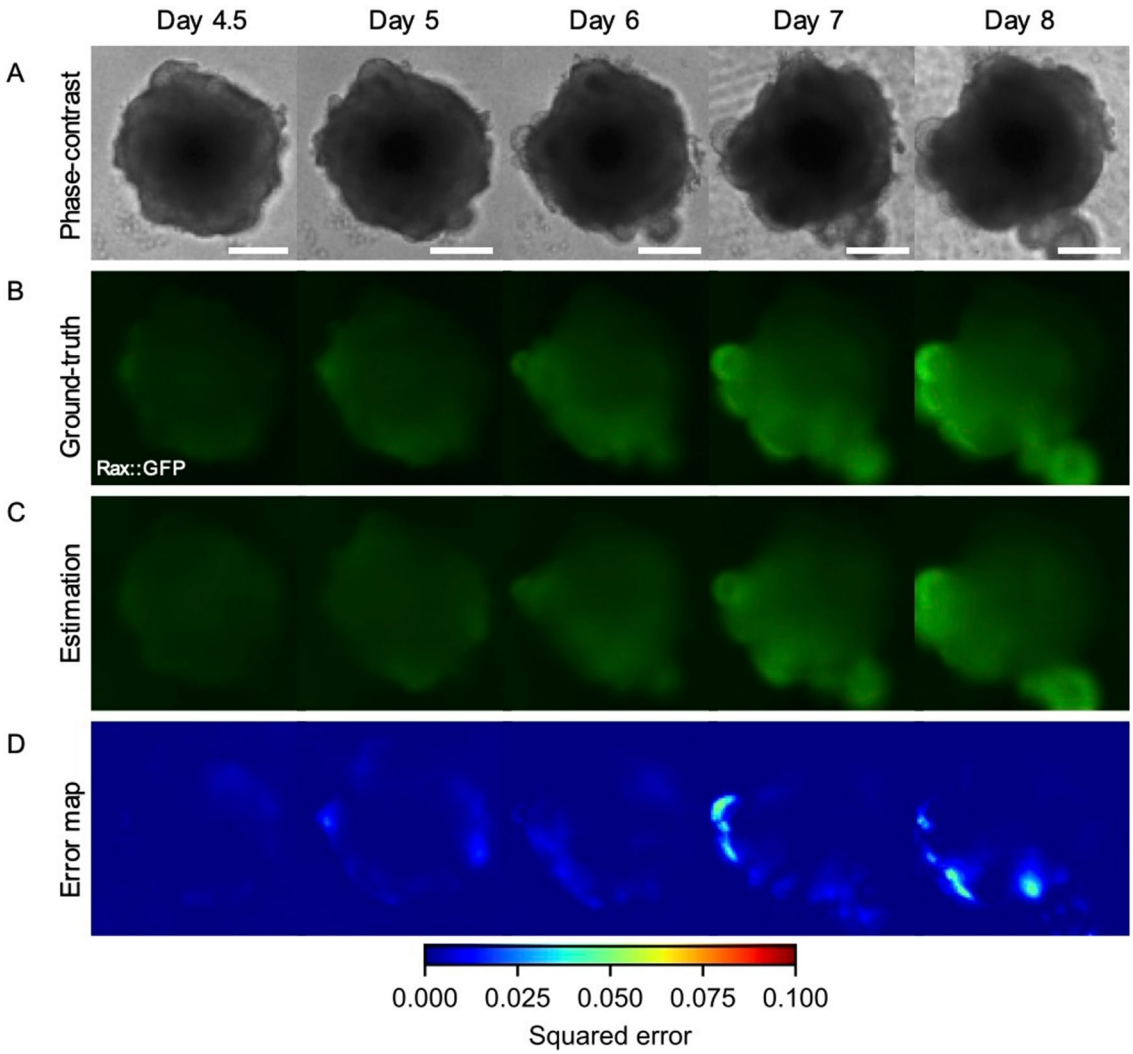

Figure 2. Estimated spatial Rax expression patterns during retinal organoid development. (**A**) Phase-contrast images from day 4.5 to day 8. (**B**) Captured fluorescent images of Rax as ground-truths. (**C**) Estimated fluorescent images with our model. (**D**) Error maps between captured and estimated images. The error metric was a squared error. The organoids in (**A**–**D**) are identical. Scale bars indicate 200 μm.

morphology of retinal organoids, especially optic vesicles, varies widely[29]. This process is known to be governed by the expression of the Rax gene (Fig. 2B). That is, the Rax gene is gradually expressed in several parts of the tissue surface, so-called eye field, where cells differentiate from neuroepithelium into several types of retinal cells.

Our model successfully recapitulated the above features of Rax expression (Fig. 2C), i.e., the Rax intensity was relatively low and homogenous at days 4.5, 5, 6, and gradually increased around the evaginated tissue regions at days 7 and 8. Remarkably, the regions of high Rax expression were accurately estimated even in organoids with various morphologies. On the other hand, as the Rax intensity increases, especially around the evaginated tissue regions, the error of the estimated image from the ground-truth image increases with time (Fig. 2D).

To quantitatively evaluate the accuracy of the estimation, we statistically analyzed the estimation results at each stage. To clarify whether the model can estimate Rax intensity in both samples with high and low Rax expression, each of the ground-truth and estimated fluorescence images was divided into two categories by the coefficient of variation of the foreground pixels in a fluorescent image at day 8 (Fig. 3A). The samples in each group were labeled as positive and negative, respectively. For each of these categories, the mean and coefficient of variation of the pixel values were calculated (Fig. 3B–E). In calculating these values, the phase-contrast images were binarized to obtain foreground and background masks, and then computed using only the foreground pixels and normalized to those of the background pixels.

Positive samples showed a gradual increase in mean and intensity over the days passed (Fig. 3B). The negative sample, on the other hand, showed relatively low values from the beginning and did not change significantly over the days (Fig. 3C). Similarly, the coefficients of variation increased in the positive samples but not in the negative samples (Fig. 3D,E). These results indicate that the model successfully estimates the feature of the spatial Rax expression patterns during retinal organoid development, i.e., positive samples gradually increase Rax expressions

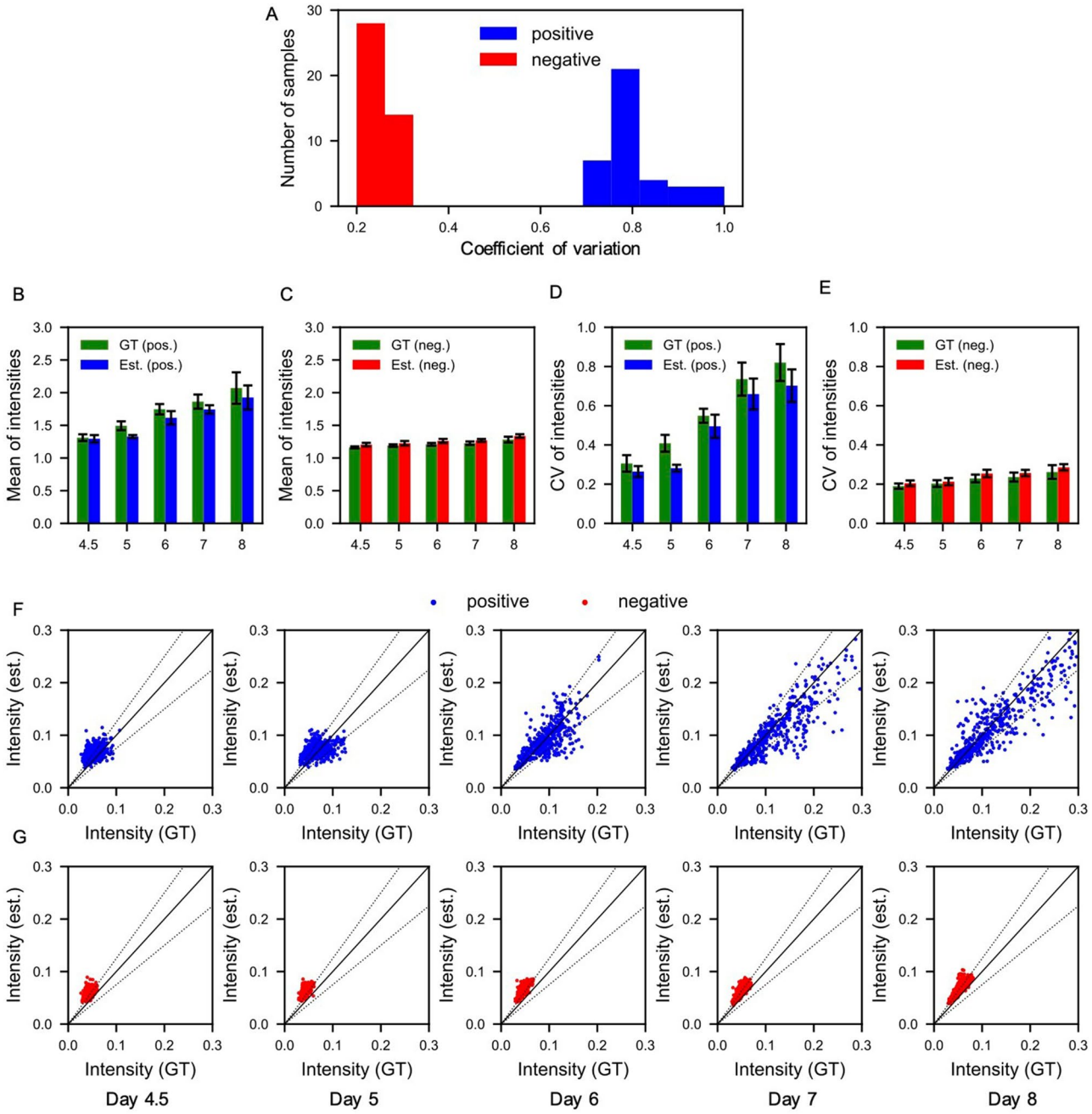

Figure 3. Statistical analysis of fluorescence at each developmental stage for positive and negative samples. (**A**) Histogram of coefficient of variation for foreground pixel values of fluorescent images at day 8. (**B**, **C**) Means of pixel values in positive and negative samples at each stage for ground-truth (green bars) and estimated fluorescent images (blue bars), respectively. (**D**, **E**) Coefficients of variation in positive samples at each stage for both ground-truth (green bars) and estimated fluorescent images (red bars), respectively. (**F**, **G**) Plots of ground-truth and estimated pixel values in positive and negative samples at each stage, respectively. Errors are 0% and 25% on the solid and dotted black lines, respectively. Error bars in (**B**–**E**) indicate standard deviations.

and their heterogeneity, but negative samples do not. The intensity of the estimated images is relatively lower than the intensity of the ground-truth images in the positive samples and vice versa in the negative samples.

To clarify whether the model is capable to estimate intermediate values of the Rax expression, we analyzed the correlations between ground-truth and estimated values on foreground pixels at each stage, respectively (Fig. 3F,G). The results show that in the positive sample (Fig. 3F), the distribution of intensities is initially concentrated at low intensities and gradually expands to high intensities as the day progresses, with a wide distribution from low to high intensities. Similarly, in the negative sample, the luminance distribution is initially concentrated at low intensities, but does not expand as much as in the positive sample (Fig. 3G). These results indicate that the model successfully estimated the plausible values across all pixel intensities, demonstrating the capability of our method to infer intermediate levels of gene expression. Notably, predicting Rax expression in

the organoids at later stages, such as day 8 in our experiments, becomes more feasible for the model due to their characteristic morphologies. These distinct morphologies provide features that can be efficiently extracted by the convolution operators of the model.

Estimated spatial Rax expression patterns correspond to organoid morphologies

To determine whether the estimated Rax expression patterns correspond to tissue morphologies, we quantified the spatial distribution of Rax intensity and the mean curvature along the tissue contour around each optic vesicle (Fig. 4). For this analysis, four typical optic vesicles were selected from the positive samples and their curvature and Rax distribution were quantified. In this analysis, tissue contours were extracted and the radius of a circle passing through three points on the tissue contour was calculated as the inverse of the curvature. Moreover, the Rax intensity was measured as the average value along the depth direction from the tissue contour.

Optic vesicles are hemispherical, with positive curvature at the distal portion and negative curvature at the root (Fig. 4A,D). The Rax intensity is continuously distributed around each vesicle, being highest at the distal part and gradually decreasing toward the root (Fig. 4B,E). Furthermore, because the test images were taken with a conventional fluorescence microscope, structures above and below the focal plane are included in each image. Therefore, although some images have multiple overlapping vesicles (e.g., samples iii and iv), the model successfully estimated the Rax intensity of the overlapping regions as well.

The best balance of the training loss weights depends on the time point

MSE is commonly used as the training loss for training regression models. In addition to MSE, this model also uses cosine similarity, which can capture structural patterns from the entire image. To analyze the effect of cosine similarity on the estimation accuracy, we tested the model with different weights for both error metrics (Fig. 5). The trained models were evaluated with MSE for each test dataset on different days (Fig. 5A). The results demonstrated that cosine similarity improved the estimation accuracy at the early and intermediate stages, such as from day 4.5 to day 6. At these stages, the intensity in the differentiated region is weak, making it difficult for the network to capture structural patterns using MSE alone. Cosine similarity, on the other hand, enabled the network to learn the patterns from the weak intensity by calculating the correlation between the normalized ground-truth and the estimated images (Fig. 5B). Therefore, our model has the capability to achieve the best estimate at different stages with appropriate weight balancing.

Discussion

To estimate spatial gene expression patterns of organoids in a non-invasive manner, we developed a machine learning-based method that converts a phase-contrast image to a fluorescent image expressing a spatial gene expression pattern. The applicability of the method was demonstrated by applying it to retinal organoids derived from mouse ESCs as an example, i.e., (i) the spatial pattern of Rax expression can be estimated during early retinal organoid development from day 4.5 to day 8, (ii) the model can be applied to both samples with positive and negative for Rax expression (retinal cs non-retinal tissues), (iii) the intermediate levels of Rax expressions can be plausibly estimated, (iv) the spatial Rax distributions can be estimated corresponding to organoid morphology , and (v) the model can be optimized to improve the accuracy by adjusting the parameters of the training loss. These results show the potential for applying machine learning to tissues defined by Rax and its orthologous genes[30] in various animals and at various stages.

A notable feature of this method is that it does not simply classify each organoid, instead, it estimates a spatial pattern of gene expression within each organoid. Most previous studies aim to apply image classification methods[31] to cell images, where the CNN takes a 2D image and outputs the probability of a semantic label[22–24,32]. The most relevant study to our research is the classification of retinal organoid images as differentiated or not[24]. On the other hand, our method employs an image-to-image translation method for cell images, such as image style transfer[33], super-resolution[34], image deburring[35], and depth (distance from a camera to an object) estimation from a monocular image[36]. The image-to-image translation method outputs a 2D image and enables pixel-wise intensity estimation. Therefore, this method is novel in that it can quantitatively estimate the spatial pattern of gene expression in each organoid.

This method can automatically and non-invasively estimate the spatial patterns of specific gene expressions in each organoid. This feature can assist in the evaluation of tissues in a wide range of biological and basic medical research. For example, this method can infer specific tissue regions without genomic alterations. Therefore, it is possible to fully analyze the endogenous gene network while maintaining the activation of the two endogenous alleles. Moreover, it avoids the ethical issues that prevent the use of genomic alterations for cell, tissue, and organ transplantation. In addition, the high throughput and automation of this method greatly reduce the time and effort required for genome editing. In addition, the method does not require human judgment, allowing for objective and robust evaluations.

In this study, the developed method was applied to Rax expression in retinal organoids as an example, while it is likely to be applicable to other organoids and other genes. In addition, the method may be applicable not only to organoids, but also to embryonic tissues, pathological tissues, and other biological tissues in principle. However, this method may not be applicable to all genes because gene expressions do not always coincide with tissue structures. For example, in retinal organoids, photoreceptors are typically located at the periphery of the retinal layer structure. It is challenging to identify the region containing photoreceptors since they are distributed both within and on the surface of the tissues. Interestingly, photoreceptor progenitors exhibit a unique behavior: they migrate within the retinal tissue through a process called interkinetic nuclear migration. Utilizing this characteristic movement, it may be possible to predict the presence of photoreceptors in tissues. Moreover, even for genes that are applicable, the model needs to be trained each time depending on the tissue type, sample

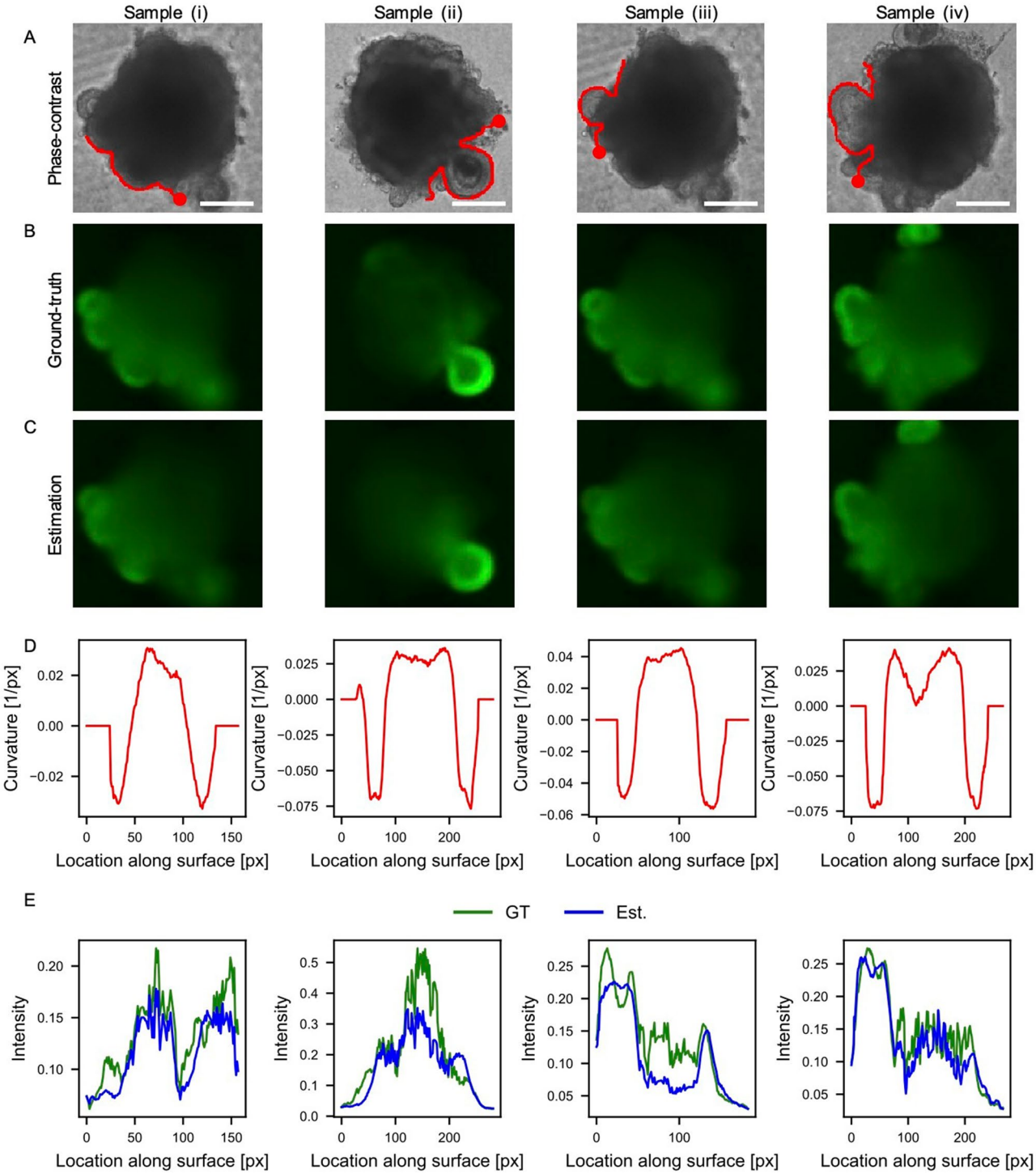

Figure 4. Correlation analysis of spatial Rax expression patterns and optic-vesicle morphologies. (**A**) Phase-contrast images. (**B**) Captured fluorescent images of Rax as ground-truths. (**C**) Estimated fluorescent images with our model. (**D**) Mean curvatures as a function of the distance along the organoid contour. (**E**) Captured and estimated fluorescent intensities of Rax along the organoid contour. The organoids in (**A**–**C**) are identical and captured on day 8. The mean curvatures and fluorescence in (**D**, **E**) are for the regions indicated by the red line starting from the red dot in (**A**). Scale bars indicate 200 μm.

preparation, and imaging method, while the estimation accuracy can be improved by adjusting the parameters according to the experimental situation (Fig. 5). Furthermore, the images used for training in this study were obtained using a standard fluorescence microscope. Therefore, while Rax expression occurs throughout the

A

(MSE, Cossim)	Day 4.5	Day 5	Day 6	Day 7	Day 8	ALL
(1.0, 0.0)	3.23 × 1e-4	4.34 × 1e-4	7.60 × 1e-4	**8.08 × 1e-4**	**11.5 × 1e-4**	**6.96 × 1e-4**
(0.01, 1.0)	**2.25 × 1e-4**	**2.88 × 1e-4**	6.43 × 1e-4	14.4 × 1e-4	20.0 × 1e-4	9.19 × 1e-4
(0.005, 1.0)	2.84 × 1e-4	3.27 × 1e-4	**5.78 × 1e-4**	9.98 × 1e-4	14.5 × 1e-4	7.28 × 1e-4
(0.001, 1.0)	13.5 × 1e-4	9.27 × 1e-4	19.5 × 1e-4	15.6 × 1e-4	18.7 × 1e-4	15.3 × 1e-4
(0.0, 1.0)	20.6 × 1e-4	14.2 × 1e-4	24.5 × 1e-4	20.3 × 1e-4	40.1 × 1e-4	23.9 × 1e-4

B

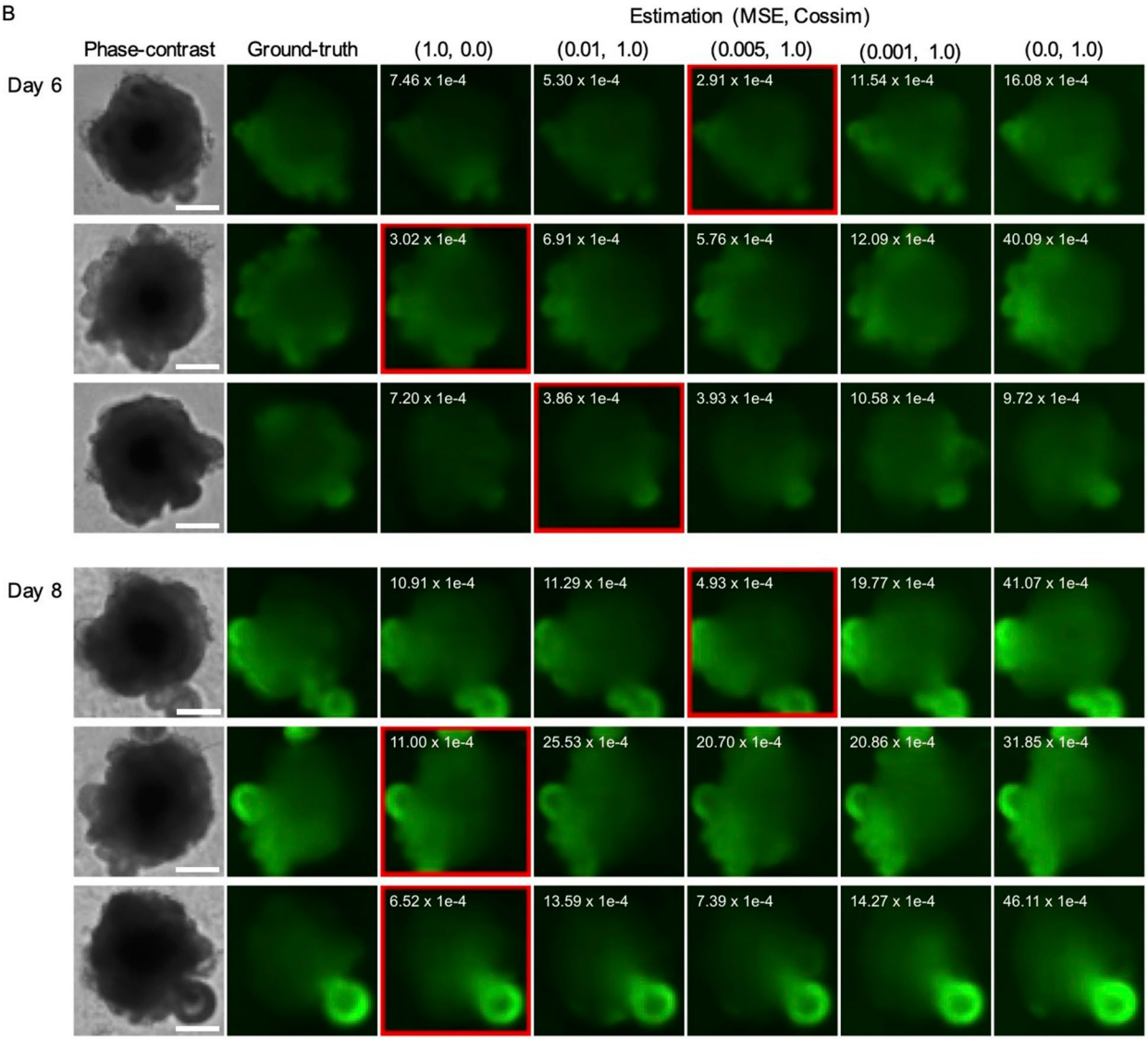

Figure 5. Effects of the balance of training loss on estimation accuracy. (**A**) Mean squared error at each stage with different hyperparameters, where bold and underlined numbers stand for the best and second best results on each day, respectively. (**B**) Examples of estimated fluorescent images at days 6 and 8 with different hyperparameters. The MSE of each estimated image is described in the upper left. The results with the lowest MSEs are surrounded by the red boxes. Scale bars indicate 200 μm.

organoid, due to the spherical shell shape of the organoid, the central region, which is thicker, is less visible under standard inverted microscopy. Estimation of such thick tissue region and further improvement of accuracy might be possible by using confocal microscope or other means. Thus, our approach opens a new avenue to assess the genetic activities of cells in organoids and other tissues in the wide range of biology and regenerative medicine.

Materials and methods

Mouse ES Rax-reporter cell line and 3D culture

Mouse ES cells (EB5, Rax-GFP) were maintained as described in the previous study[10,37]. The cell line used in this study is a subline of the mouse ESC line, EB5 (129/Ola), in which the GFP gene was knocked-in under the Rax promoter. The cell line was obtained from the RIKEN BioResource Research Center (ID: AES0145).

Cells were maintained as described in the previous study[10]. For maintenance, cells were cultured in a gelatin-coated 10 cm dish. The dish contained maintenance medium, to which 20 µl of 10^6 units/ml LIF (Sigma-Aldrich), and 20 µl of 10 mg/ml blastacidin (Cayman) were added. Cells were incubated at 37 °C in 5% CO_2. The maintenance medium consisted of Glasgow's Modified Eagle Medium (G-MEM; Wako) supplemented with 10% Knockout Serum Replacement (KSR; GIBCO), 1% Fetal Bovine Serum (FBS; GIBCO), 1% Non-essential Amino Acids (NEAA; Wako), 1 mM pyruvate (Wako), and 0.1 mM 2-Mercaptoethanol (2-ME; Sigma-Aldrich). The solution was filtered through a 0.2 µm filter bottle, stored at 4 °C and used within one month.

The SFEBq culture method was performed as described in the previous study[10]. In this method, 3000 cells were suspended in 100 µl of differentiation medium in each well of a 96-well plate, marking day 0. On day 1, Matrigel (Corning) was mixed with the differentiation medium and added to each well to reach a final concentration of 2.0%. This plate was incubated at 37 °C in a 5% CO_2 environment until day 8. The maintenance medium consisted of G-MEM supplemented with 10% KSR, 1% FBS, 1% NEAA, 1 mM pyruvate, and 0.1 mM 2-ME. This solution was filtered through a 0.2 µm filter bottle, stored at 4 °C, and used within one month. 96 organoids were used in this study.

Automated imaging

Both phase-contrast and fluorescent images were captured using a Keyence All-In-One Fluorescence Microscope BZ-X800. A 10X objective lens was used. On each day from 4.5 to 8 during SFEBq culture, 14 images were taken for each organoid at different Z positions with 15 µm intervals within 195 µm. The resolution of both phase-contrast and fluorescent images is 960 × 720, from which the 512 × 512 regions where organoids appear were extracted; phase-contrast images were binarized and the median pixels in the vertical and horizontal axes were computed, and then the 512 × 512 regions centered at these median pixels were cropped.

Quantification of curvature and Rax distributions

To quantify tissue curvature and Rax distributions, tissue contours were first manually extracted from the phase-contrast images. Then, the radius of a circle passing through three points on the tissue contour (the point where the curvature is measured and two points 81.6 µm before and after that point) was calculated and its reciprocal was defined as the curvature. The average of the intensity from the point at which the Rax expression intensity is measured on the tissue contour to 13.6 µm vertically from the tangent line of that point was also measured.

Training loss

The training loss $\mathcal{L}$ is computed from the output fluorescent images and the corresponding ground-truth images. The trainable parameters θ (the weights of these convolution kernels and biases for CNNs) are optimized by minimizing the training loss:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\ \mathbb{E}_{\{\mathbf{x},\mathbf{y}\}\sim\mathcal{D}}[\mathcal{L}],$$

where $\mathcal{D}$ is the training dataset, and $\mathbf{x}$ and $\mathbf{y}$ are a phase-contrast and the corresponding fluorescent image, respectively. A naïve loss function is the mean squared error (MSE), where the training loss is computed pixel by pixel as follows:

$$\mathcal{L}_{\text{mse}} = \frac{1}{N'}\sum_{i=1}^{N'}\left(\mathbf{y}_{i'} - \widehat{\mathbf{y}}_i\right)^2,$$

where $\mathbf{y}_{i'}$ and $\widehat{\mathbf{y}}_i$ are the values at the i-th pixel in the ground-truth and estimated images, respectively. N' is a number of pixels. In this study, we additionally use the cosine similarity as follows:

$$\mathcal{L}_{\text{cossim}} = -\frac{\mathbf{y}'^{\top}\widehat{\mathbf{y}}}{\|\mathbf{y}'\|_2\|\widehat{\mathbf{y}}\|_2}.$$

This makes the network learn a global structure, because the loss is computed considering the whole pixels simultaneously. We combine the above two losses as follows:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{mse}} + \beta\mathcal{L}_{\text{cossim}},$$

where α and β are hyperparameters to control the contribution of each loss.

Implementation details

Our network was implemented in PyTorch. Training was performed on an NVIDIA A100 GPU. The size of a mini-batch was 16 for the training loss. The encoder was initialized by the ResNet50 pre-trained on the ImageNet[38]. The optimizer was Adam[39] with the default hyperparameters ($lr = 10^{-3}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$). The learning rate was multiplied by 0.1 at every 20 epochs. The network was trained for 50 epochs and we adopted the best model on the validation dataset.

Data availability

The trained model, sample code and sample data used for running experiments are available in a GitHub repository at https://github.com/yfujimura/deep-organoid.

Received: 11 May 2023; Accepted: 12 December 2023
Published online: 20 December 2023

References

1. Zhao, Z. *et al.* Organoids. *Nat. Rev. Methods Primers* **2**, 94 (2022).
2. Grassi, L. *et al.* Organoids as a new model for improving regenerative medicine and cancer personalized therapy in renal diseases. *Cell Death Dis.* **10**, 201 (2019).
3. Nakamura, T. & Sato, T. Advancing intestinal organoid technology toward regenerative medicine. *Cell Mol. Gastroenterol. Hepatol.* **5**, 51–60 (2018).
4. Takebe, T. & Wells, J. M. Organoids by design. *Science* **1979**(364), 956–959 (2019).
5. Perez-Gonzalez, C., Ceada, G., Matejčić, M. & Trepat, X. Digesting the mechanobiology of the intestinal epithelium. *Curr. Opin. Genet. Dev.* **72**, 82–90 (2022).
6. Nath, S., Toda, S. & Okuda, S. Intestinal and optic-cup organoids as tools for unveiling mechanics of self-organizing morphogenesis. *Biophys. Physicobiol.* **19**, e190048 (2022).
7. Sasai, Y. Cytosystems dynamics in self-organization of tissue architecture. *Nature* **493**, 318–326 (2013).
8. Madrid, M., Sumen, C., Aivio, S. & Saklayen, N. Autologous induced pluripotent stem cell–based cell therapies: Promise, progress, and challenges. *Curr. Protoc.* **1**, e88 (2021).
9. Mandai, M. Pluripotent stem cell-derived retinal organoid/cells for retinal regeneration therapies: A review. *Regen. Ther.* **22**, 59–67 (2023).
10. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–56 (2011).
11. Hasegawa, Y. *et al.* Emergence of dorsal-ventral polarity in ESC-derived retinal tissue. *Development* **143**, 3895–3906 (2016).
12. Okuda, S. *et al.* Strain-triggered mechanical feedback in self-organizing optic-cup morphogenesis. *Sci. Adv.* **4**, eaau1354 (2018).
13. Takata, N., Sakakura, E., Eiraku, M., Kasukawa, T. & Sasai, Y. Self-patterning of rostral-caudal neuroectoderm requires dual role of Fgf signaling for localized Wnt antagonism. *Nat. Commun.* **8**, 1339 (2017).
14. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
15. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
16. Beghin, A. *et al.* Automated high-speed 3D imaging of organoid cultures with multi-scale phenotypic quantification. *Nat. Methods* **19**, 881–892 (2022).
17. Schuster, B. *et al.* Automated microfluidic platform for dynamic and combinatorial drug screening of tumor organoids. *Nat. Commun.* **11**, 5271 (2020).
18. Driehuis, E., Kretzschmar, K. & Clevers, H. Establishment of patient-derived cancer organoids for drug-screening applications. *Nat. Protoc.* **15**, 3380–3409 (2020).
19. Takata, N. & Eiraku, M. Stem cells and genome editing: Approaches to tissue regeneration and regenerative medicine. *J. Hum. Genet.* **63**, 165–178 (2018).
20. Artegiani, B. *et al.* Fast and efficient generation of knock-in human organoids using homology-independent CRISPR–Cas9 precision genome editing. *Nat. Cell Biol.* **22**, 321–331 (2020).
21. Coronnello, C. & Francipane, M. G. Moving towards induced pluripotent stem cell-based therapies with artificial intelligence and machine learning. *Stem Cell Rev. Rep.* 1–11 (2022).
22. Zhu, Y. *et al.* Deep learning-based predictive identification of neural stem cell differentiation. *Nat. Commun.* **12**, 2614 (2021).
23. Waisman, A. *et al.* Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation. *Stem Cell Rep.* **12**, 845–859 (2019).
24. Kegeles, E., Naumov, A., Karpulevich, E. A., Volchkov, P. & Baranov, P. Convolutional neural networks can predict retinal differentiation in retinal organoids. *Front. Cell Neurosci.* **14**, 171 (2020).
25. Gonzalez-Cordero, A. *et al.* Photoreceptor precursors derived from three-dimensional embryonic stem cell cultures integrate and mature within adult degenerate retina. *Nat. Biotechnol.* **31**, 741–747 (2013).
26. Chen, H. Y., Kaya, K. D., Dong, L. & Swaroop, A. Three-dimensional retinal organoids from mouse pluripotent stem cells mimic in vivo development with enhanced stratification and rod photoreceptor differentiation. *Mol. Vis.* **22**, 1077 (2016).
27. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* 234–241 (2015).
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
29. Decembrini, S., Koch, U., Radtke, F., Moulin, A. & Arsenijevic, Y. Derivation of traceable and transplantable photoreceptors from mouse embryonic stem cells. *Stem Cell Rep.* **2**, 853–865 (2014).
30. Bailey, T. J. *et al.* Regulation of vertebrate eye development by Rx genes. *Int. J. Dev. Biol.* **48**, 761–770 (2004).
31. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. in *Advances in Neural Information Processing Systems* (eds. Pereira, F., Burges, C. J., Bottou, L. & Weinberger, K. Q.) vol. 25 (Curran Associates, Inc., 2012).
32. Kusumoto, D. *et al.* Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. *Stem Cell Rep.* **10**, 1687–1695 (2018).
33. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1125–1134 (2017).
34. Ledig, C. *et al.* Photo-realistic single image super-resolution using a generative adversarial network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4681–4690 (2017).
35. Dong, J., Roth, S. & Schiele, B. Deep wiener deconvolution: Wiener meets deep learning for image deblurring. *Adv. Neural Inf. Process Syst.* **33**, 1048–1059 (2020).
36. Hu, J., Ozay, M., Zhang, Y. & Okatani, T. Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries. in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1043–1051 (2019).
37. Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nat. Neurosci.* **8**, 288–296 (2005).
38. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (2009).
39. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *Proceedings of the 3rd international conference for learning representations (ICLR '15)* **500**, (2015).

Acknowledgements
This work was supported by the Japan Science and Technology Agency (JST), CREST Grant No. JPMJCR1921 (S.O.), JPMJCR23N3 (T.F.) and PRESTO Grant No. JPMJPR2025 (T.F.); the Japan Agency for Medical Research and Development (AMED), Grant No. 21bm0704065h0003 (S.O.); the Japan Society for the Promotion of Science (JSPS), KAKENHI Grants No. 21H01209, 21KK0134, 22K18749, and 22H05170 (S.O.); and the World Premier International Research Center Initiative, Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (S.O).

Author contributions
Y.F., A.H., T.F. and S.O. conceptualized the project. Y.F., A.H. and T.F. developed the methodology. Y.F. created the software used. I.S. and I.S. curated the data and conducted the investigation. S.O. acquired funding, administered the project, provided resources. Y.F. and S.O. wrote the original draft of the manuscript. All authors reviewed and edited the manuscript, and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.F. or S.O.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023