APPLIED SCIENCES AND ENGINEERING

Deep manifold learning reveals hidden developmental dynamics of a human embryo model

Kejie Chen[1,2]*, Kai-Rong Qin[3], Jing Na[1,2], Guanbin Gao[1,2], Chunxi Yang[1,2], Jianping Fu[4,5,6]*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

In this study, postimplantation human epiblast and amnion development are modeled using a stem cell–based embryoid system. A dataset of 3697 fluorescent images, along with tissue, cavity, and cell masks, is generated from experimental data. A computational pipeline analyzes morphological and marker expression features, revealing key developmental processes such as tissue growth, cavity expansion, and cell differentiation. To uncover hidden developmental dynamics, a deep manifold learning framework is introduced. This framework uses an autoencoder to project embryoid images into a twenty-dimensional (20D) latent space and models the dynamics using a mean-reverting stochastic process of mixed Gaussians. The approach accurately captures phenotypic changes observed at discrete experimental time points. Moreover, it enables the generation of artificial yet realistic embryoid images at finer temporal resolutions, providing deeper insights into the progression of early human development.

INTRODUCTION

Understanding human development is important for advancing human reproductive medicine and promoting regenerative medicine applications (*1*, *2*). However, because of limited access to human embryonic tissues and ethical and technical challenges, regulatory mechanisms of cell lineage diversification and morphogenetic changes during early human development still remain largely unknown (*3*, *4*). Recently, human pluripotent stem cells (hPSCs) have shown the potential of self-organizing into embryonic-like structures and recapitulating key embryonic cell fate decision and tissue patterning events. These embryo models (also known as embryoids), although still far from perfectly mimicking in vivo embryonic development, provide exciting opportunities to study signaling dynamics, cell fate decision, tissue patterning, and morphogenesis in the context of human development at a high spatiotemporal resolution (*5*–*9*).

The molecular and cellular features and the transitions between developmental stages of the human embryoid development are commonly characterized via bioimaging with fluorescent microscopies. On the basis of lineage marker expression profiles, morphological features, and cellular organization, these human embryoids are then benchmarked against well staged human embryo references [e.g., the Carnegie stages (*10*)]. However, human embryoids in a culture dish often exhibit asynchronous development, resulting in pronounced heterogenous phenotypic features. The transitions between developmental stages in a human embryoid occur progressively, which is therefore difficult to strictly define a sharp developmental transition (*11*, *12*). In addition, many other practical experimental factors—such as limited temporal resolution of sampling, small or unavailable dataset, and tedious and subjective identification by human eyes—limit our understanding of the developmental dynamics of human embryoids.

Computational methods capable of automatic image profiling have been used to analyze bioimages and offer an efficient way to understand the molecular, cellular, and morphogenetic features of biological tissues (*13*–*19*). These bioimaging tools, including CellProfiler (*16*, *17*), TrackMate (*18*), and Ilastik (*19*), use both conventional computer vision algorithms and deep learning models for processing large and complex bioimage datasets. However, until now, these approaches have rarely been applied for bioimaging of multicellular systems including human embryoids. Because, unlike conventional images of cell cultures on two-dimensional (2D) tissue culture plates, image-specific, label-intensive, manual adjustments, and annotations of embryoid images are often needed because of their complex 3D structures and lower image quality. The acquisition of adequate amount of high-quality embryoid images remains challenging, given the low efficiencies of current embryoid development protocols, their long-term culture, and time-consuming immunostaining and imaging procedures.

Nonetheless, a few recent studies have achieved successful phenotypic screens and morphological quantification of human organ models or organoids derived from stem cells, whose development and culture protocols have been well developed (*20*–*24*). For example, Lukonin *et al.* (*23*) generated about half a million of intestinal organoids under the treatments of a library of 2789 chemical compounds and analyzed multivariate organoid features using a custom image acquisition and processing pipeline. Betge *et al.* (*24*) acquired more than 3 million confocal images of patient-derived colorectal cancer organoids treated with more than 500 small molecules. They applied ResNet model and graph-based clustering to study the phenotypic landscapes of cancer organoids.

The efficiency, throughput, cost, and associated ethical issues of human embryoids are much more challenging compared to those associated with cancer and intestinal organoids. Another compounding issue is that there are very few datasets of early postimplantation human and primate monkey embryos and their corresponding immunostaining images (*25*, *26*). Recently, we have successfully cultured more than 3000 hPSC-based embryoids, which recapitulate postimplantation human epiblast and amnion development and the formation of the fluid-filled amniotic cavity, and acquired their immunostaining images after optimizing the culture protocol (*12*). Several other protocols of generating

[1]Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, Yunnan, China. [2]Yunnan Key Laboratory of Intelligent Control and Application, Kunming, Yunnan, China. [3]School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, Liaoning, China. [4]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA. [5]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA. [6]Department of Cell & Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA.
*Corresponding author. Email: kejie_2020@163.com (K.C.); jpfu@umich.edu (J.F.)

mouse (*27*, *28*) and human embryoids (*5*–*9*, *29*) have achieved a throughput of a few thousand samples as well.

On the basis of experimentally collected images of human embryoids, we developed a custom deep learning image processing pipeline to generate a clean, high-quality dataset capturing progressive human embryoid development. This dataset comprises 3697 fluorescent images of human embryoids immunostained for various cell lineage markers, along with segmentation masks of embryoid tissues and individual cells associated with each image. Our pipeline includes a deep manifold learning framework integrated with stochastic dynamics modeling. Morphological and fluorescence features of each fluorescent image were first analyzed to quantitatively characterize key developmental features of human embryoids, including growth of embryoid tissues, emergence and expansion of a fluid-filled, amniotic-like cavity, and lineage-specific cell differentiation. To further investigate spatial cell organization, we applied a graph-based method to quantify local cell distributions and cell-cell connectivity networks, revealing evidence of cell type–specific sorting within human embryoids. We also used an encoder-decoder neural network to project fluorescent images into a low-dimensional latent feature space. This analysis showed that five latent dimensions were sufficient to capture essential morphological features of human embryoids, while 20 dimensions effectively represented lineage marker expression patterns in the human embryoids. Within this latent space, the temporal evolution of feature distributions was robustly modeled using a mean-reverting stochastic process with mixed Gaussian components. Our results indicate that two time-varying Gaussian distributions adequately approximate the experimental data. Last, by integrating the learned stochastic dynamics with the encoder-decoder network, we generated artificial but realistic embryoid images at finer temporal resolution, enabling the reconstruction of continuous developmental trajectories of human embryoids from a limited number of experimentally observed time points.

RESULTS

The deep manifold learning framework for analyzing human embryoid development consists of three major components, as illustrated in Fig. 1. First, an embryoid dataset was constructed, comprising aligned raw images and corresponding masks of tissues and cells. Second, morphological and lineage marker intensity features of human embryoids were characterized, with particular emphasis on quantifying their temporal changes and spatial distributions, and the structure of cell-cell connectivity networks. Last, a stochastic dynamic-integrated encoder-decoder model was developed to extract low-dimensional, whole image–based latent features, to generate artificial yet realistic sample images, and to infer the continuous-time regulatory dynamics underlying tissue development and morphologic dynamics in human embryoids.

Embryoid image dataset

hPSC-based embryoids were generated using a protocol described in our previous studies (*12*). Singly dissociated hPSCs were first aggregated

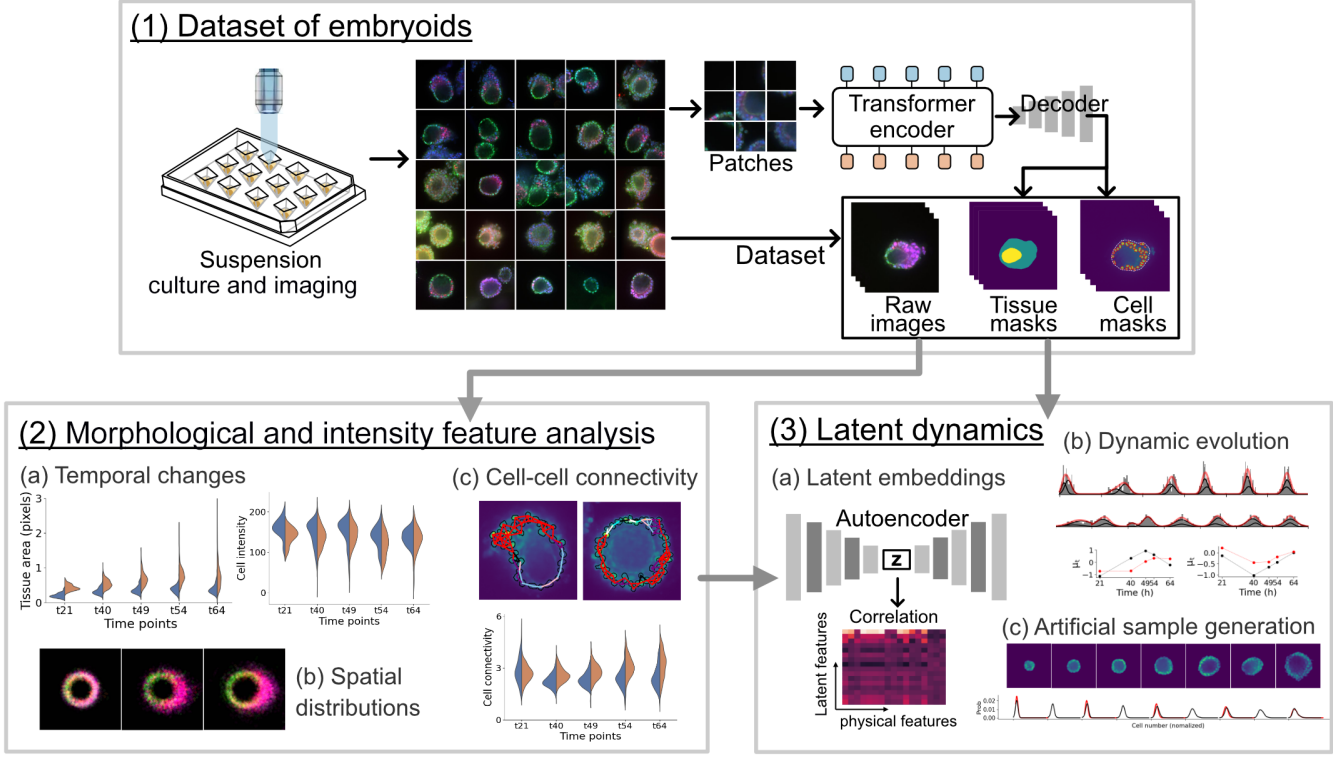


Fig. 1. Deep manifold learning framework for characterizing developmental dynamics of human embryoids. (1) Establishment of a rich image dataset of more than three thousand embryoids. Fluorescent images of human stem cell–based embryoids were acquired at multiple time points over a period of 64 hours. A custom image processing pipeline was applied for image alignment, size adjustment, and segmentation of embryonic-like tissues and single cells. (2) Quantification of morphological and molecular features. (3) Study of continuous, temporal developmental dynamics in the latent feature space, extracted using an autoencoder neural network. The mean-reverting diffusion dynamics of mixed Gaussian distributions were applied to model underlying regulatory dynamics of progressive embryoid development.

in pyramidal microwells at three different initial seeding densities, resulting in ~20 (group 1), 50 (group 2), or 80 (group 3) cells per well. After 6 hours of culture in the pyramidal wells, the aggregates were transferred to 3D Matrigel and cultured for an additional 64 hours (Fig. 2A). At five time points between $t = 12$ to 60 hour, a subset of embryoids was collected, immunostained for lineage markers (NANOG, GATA3, and BRACHYURY or T), and imaged. Specifically, NANOG was used to identify epiblast-like cells, while GATA3 and T marked amnion- and mesoderm-like cells, respectively (*6*, *8*, *12*).

Embryoid images were pre-processed using a custom image processing pipeline. As illustrated in Fig. 2B, the pipeline first centered each embryoid within the image, followed by cropping and resizing the image to 288 × 288 pixels. Next, embryonic-like tissues, fluid-filled cavities, and individual cells within the tissues were segmented using transformer–convolutional neural networks (CNNs). The transformer-CNN architecture (Fig. 2C), previously reported in (*30*), was adopted for image segmentation, with the transformer head modified to accept four-channel input images. To train the networks, images of human embryoids in group 1 and group 2 were manually annotated to provide ground truths for embryonic tissues and fluid-filled cavities. In addition, 1000 images of human embryoids in group 2 were annotated for single-cell segmentation. The annotated images were then used for network training and validation. Further details are provided in the "Image processing pipeline" section in Materials and Methods. Figure 2D presents examples of raw cropped images, manual annotations, and the corresponding predictions by the transformer-CNN networks. Training of five independently initialized networks converged within 1000 iterations, achieving both cross-entropy (CE) loss and dice loss below 0.01 (Fig. 2, E and F). Compared to manual annotations, the network-generated masks were generally less subjective and featured smoother boundaries. The trained

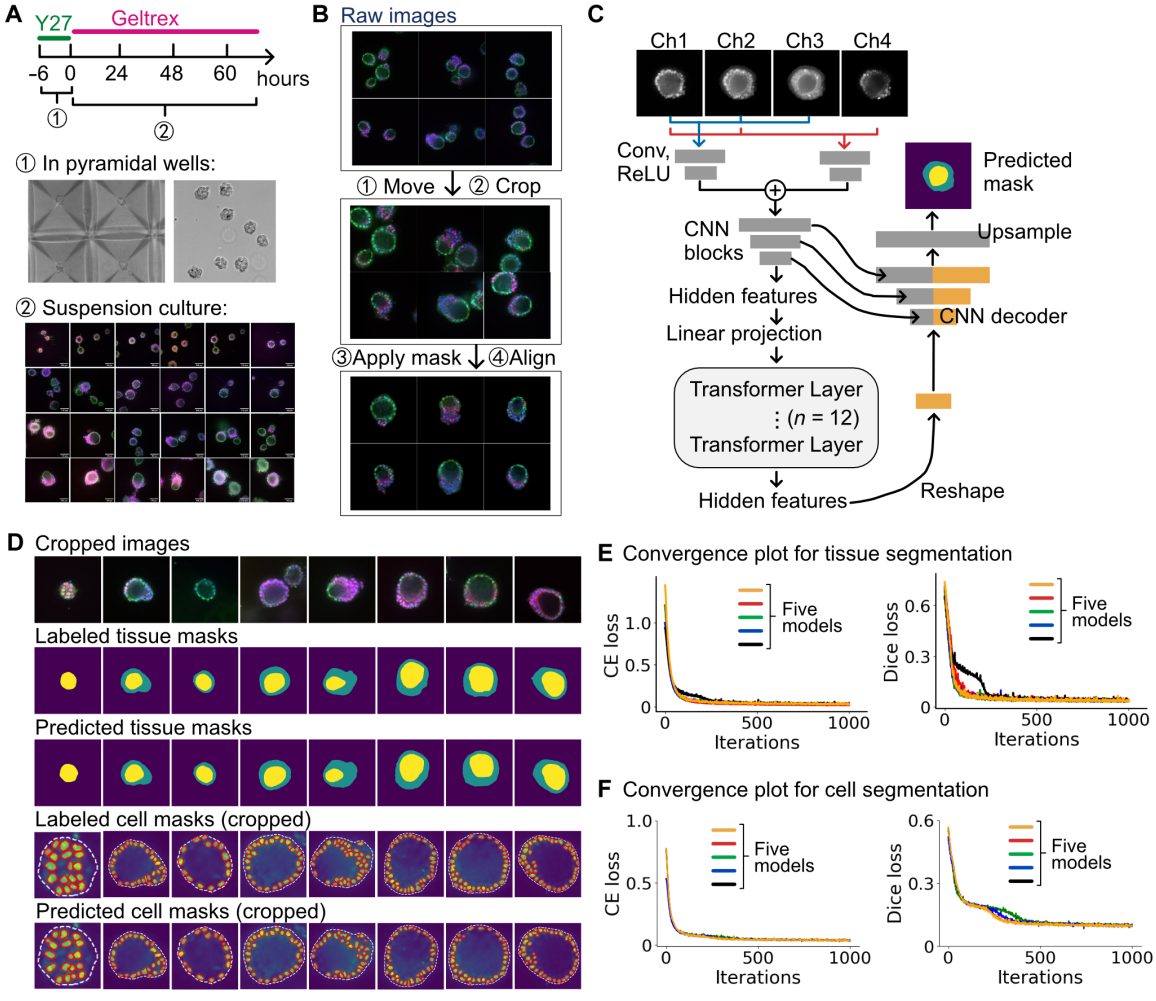


Fig. 2. Establishment of the embryoid image dataset. (**A**) Fabrication and imaging of the embryoids. Embryoids were fabricated by aggregating human embryonic stem cells (hESCs) in the pyramidal wells and cultured in the 3D gel matrix. Samples collected at multiple time points were immune-stained by fluorescent makers of DAPI, NANOG, GATA3, and T and imaged using confocal microscopy. (**B**) Self-devised image processing pipeline, which contains size adjustment, segmentation, alignment, and standardization. (**C**) Architecture of the transformer-CNN network for image segmentation. (**D**) Examples of experimentally acquired raw images, manually labeled masks and predicted masks of the embryonic tissues, cavities, and single cells by the transformer-CNN network. Green color is for embryonic tissues, yellow is for central cavity, red boundaries indicate single cells, and white dashed curves show the outer boundaries of the embryoids. (**E**) Convergence plots of the cross-entropy (CE) losses and dice losses for tissue segmentation. (**F**) Convergence plots of the CE losses and dice losses for segmentation of single cells.

transformer-CNN networks were subsequently applied to generate masks for the entire embryoid dataset.

After preprocessing, a total of 3697 images were collected. These images were divided into three groups corresponding to small, intermediate, and large initial cell seeding densities, containing 795, 1424, and 1478 images, respectively. Each image contained a single segmented embryoid centered and vertically aligned. Two associated masks, one for embryonic-like tissues and fluid-filled cavities and another for single cells, were generated and stored with each image. Example raw images acquired from the experiments are shown in figs. S1 and S2, while figs. S3 and S4 present examples of postprocessed images. In addition, the performance of the transformer-CNN model was compared with the watershed method used in previous work (*12*), as shown in fig. S5. While both methods produced similar boundaries for embryoid tissues and cavities, the transformer-CNN model achieved significantly higher accuracy in segmenting single cells.

Phenotypic features of embryoids

Morphological properties of the embryoids—including tissue area, cavity area, tissue thickness ratio, eccentricity, perimeter, and major/minor axis lengths—were quantified on the basis of tissue and cavity masks. Cell-wise fluorescent features were extracted using single-cell masks, including the total number of cells per image slice; the numbers of $GATA3^+$, $NANOG^+$, and T^+ cells; and the average fluorescence intensity of each marker at the single-cell level. Definitions of these quantitative phenotypic features are provided in table S1. A schematic illustration of the tissue thickness ratio, axis lengths, marker-positive cells, and cell-cell connectivity is shown in fig. S6.

A custom Python algorithm pipeline was developed to automatically identify marker-positive cells and calculate both their morphological and fluorescent features. To visualize cell distributions, images containing only marker-positive cells were aligned by positioning the tissue region with the smallest thickness on the left, followed by generating superimposed images of all marker-positive cells (Fig. 3A). In these super-impositions, regions with many overlapping cells appear brighter due to higher pixel-wise color intensity. Therefore, these composite images serve as probabilistic maps of the spatial distributions of $GATA3^+$, $NANOG^+$, and T^+ cells in the human embryoids (Fig. 3B).

As shown in Fig. 3B, all embryoids exhibited a symmetric and spherical shape, with central spherical fluid-filled cavities, when $t < 40$ hours of culture. After $t = 48$ hours, many human embryoids became elongated, and their tissue thickness ratios increased significantly. On the basis of these morphological changes, the human embryoids were classified as either symmetric or asymmetric using a threshold tissue thickness ratio of 1.5. After $t = 40$ hours, $GATA3^+$ cells began to emerge on the thinner side of the human embryoid. In group 1 from the least initial cell seeding density, the appearance of $GATA3^+$ cells was delayed until $t \approx 54$ hours. Both symmetric and asymmetric embryoids contained $GATA3^+$ and T^+ cells, which became spatially segregated to opposite sides of the embryoid tissue after prolonged culture. This observation is consistent with a previous study (*29*), which reported that anterior-posterior symmetry was broken in an hPSC-based epiblast model cultured in 3D Matrigel under a low dose of BMP4 treatment. Notably, around $t = 54$ hours in group 1 and $t = 48$ to 49 hours in groups 2 and 3, $GATA3^+$ cells were present in the human embryoids but appeared randomly distributed and intermingled with other cell types. A clear spatial segregation of $GATA3^+$ cells within symmetric tissues was observed only when $t > 60$ hours, whereas this sorting of $GATA3^+$ cells occurred earlier in asymmetric embryoids.

We quantified the average fluorescent intensity of $GATA3^+$, T^+, and $NANOG^+$ cells across the three human embryoid groups. GATA3 intensity was higher at early culture times ($t < 40$ hours; Fig. 3, C to E), although only very few $GATA3^+$ cells were present at this early stage (Fig. 3F). T intensity peaked at $t \approx 49$ hours in groups 1 and 2 and around $t \approx 24$ hours in group 3 (Fig. 3, C to E), while the number of T^+ cells showed only slight fluctuations over time (Fig. 3F). NANOG intensity decreased after $t > 49$ and 36 hours in groups 1/2 and 3, respectively (Fig. 3, C to E). Notably, the tissue thickness surrounding $GATA3^+$ cells significantly decreased in the human embryoids only after $t > 64$ hours. The increase in $GATA3^+$ cell number (Fig. 3F) and the decrease in tissue thickness (Fig. 3G) suggest that the morphological change of the amnion-like tissues and $GATA3^+$ cell accumulation occurred much later than the initial expression of GATA3. Last, Fig. 3 (G and H) shows that some human embryoids in groups 2 and 3 grew to a very large size, with more than 50 cells present in a single image slice. In these cases, tissue area and total cell number no longer scaled proportionally with the initial cell seeding density.

Cell-cell connectivity networks were constructed on the basis of the segmentation results of single cells. Two cells were considered connected, and thus potentially able to communicate, when the distance between their centers was less than a threshold (20 pixels in this study). Example connectivity networks are shown in Fig. 4A. As illustrated in Fig. 4B, the average cell connectivity was approximately three at early culture time points across all sample groups, meaning each cell was connected to about three neighboring cells. In group 1, cell connectivity decreased slightly over time, possibly due to the expansion of the central cavity. In groups 2 and 3, connectivity only increased after $t > 54$ and 48 hours, respectively (Fig. 4B), despite the continuous increase in total cell numbers throughout the 54-hour culture period (Fig. 3I). Fluorescent images show that in early-stage symmetric human embryoids, cells formed only one or two layers surrounding the central cavity. During this stage, cell proliferation primarily contributed to the enlargement of both the tissue and the central cavity, without significantly altering cell connectivity. In the later stages (after $t > 48$ hours), T^+ cells began to accumulate on one side of the embryoid tissue, leading to a rise in local cell connectivity. For $t < 40$ hours for group 1 and $t < 36$ hours for groups 2 and 3, the connectivity between different cell types remained around 0.5 (Fig. 4C), indicating a salt-and-pepper distribution of different cell populations. After $t > 40$ hours, this connectivity dropped significantly in groups 2 and 3 (Fig. 4C), suggesting active cell migration and sorting. As a result, $GATA3^+$ and T^+ cells became spatially segregated and accumulated at opposite sides of the embryoid tissues (Fig. 3B). The connectivity between $GATA3^+$ and T^+ cells peaked between $t = 36 - 49$ hours and then decreased after $t > 49$ hours (Fig. 4D). The initial increase in the connectivity between $GATA3^+$ and T^+ cells likely resulted from continuous cell differentiation into these two lineages in the human embryoids, while the subsequent decline in the connectivity between them reflected the effects of cell sorting.

Latent feature embeddings

Beyond visible phenotypic features, many properties of human embryoids are imperceptible to the human eye or cannot be easily

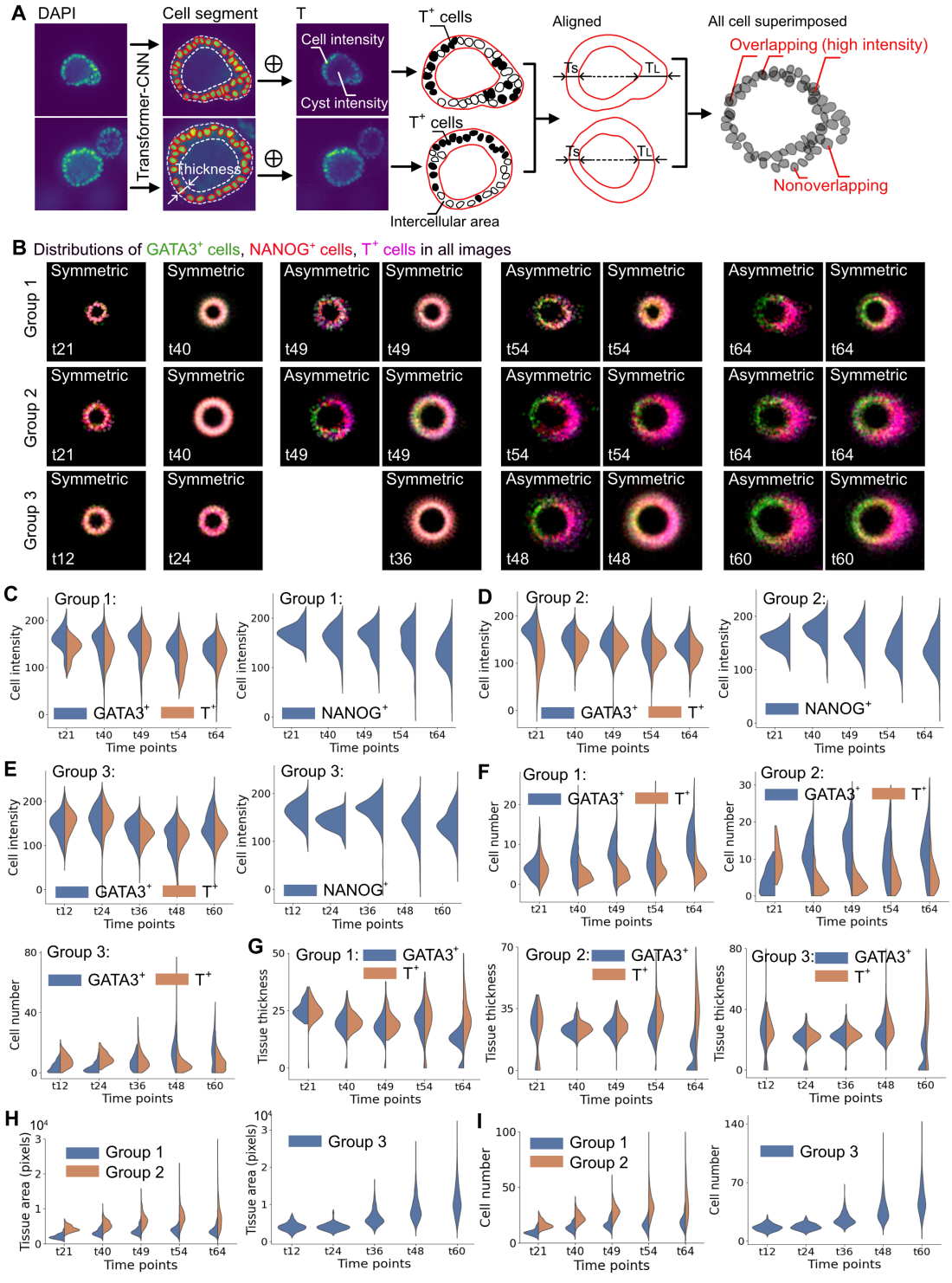


Fig. 3. Morphological and cellular fluorescent properties of embryoid images in the dataset. (**A**) Feature analysis pipeline includes segmentation of embryonic tissues and single cells; identification of GATA3$^+$, NANOG$^+$, and T$^+$ cells; measurement of morphological and fluorescent intensity features; and visualization of the cell distributions. (**B**) Probabilistic distributions of GATA3$^+$ cells, NANOG$^+$ cells, and T$^+$ cells at five culture time points in the three groups grown from small, intermediate, and large initial cell densities. Color represents three marker types. Large color brightness indicates that there are more marker-positive cells. An embryoid is classified as an asymmetric tissue if its thickness ratio is larger than 1.5. Otherwise, it is regarded as a symmetric tissue. (**C** to **E**) Group-wise distributions of mean cellular intensities of GATA3, NANOG, and T markers over time. (**F**) Group-wise distributions of the number of GATA3$^+$, NANOG$^+$, and T$^+$ cells. (**G**) Group-wise distributions of tissue thickness surrounding the GATA3$^+$, NANOG$^+$, and T$^+$ cells. (**H**) Group-wise distributions of tissue area over time. (**I**) Group-wise distributions of cell number over time.

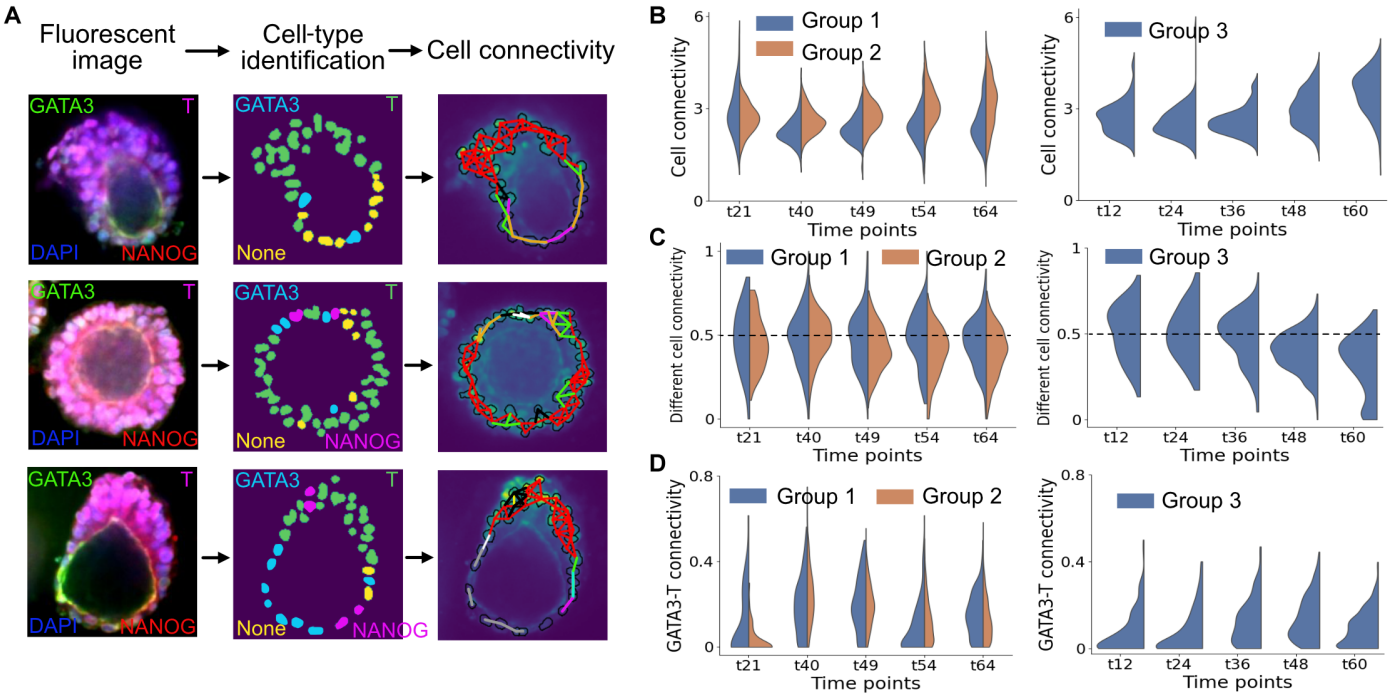


Fig. 4. Cell-cell connectivity network showing that cell-type–based sorting happens during the emrbyoid development. (**A**) Establishment of the cell-cell connectivity map. If the distance between two cells are smaller than a threshold (i.e., 20 pixels in this study), then the cell-cell connection is assumed be established, which is represented by a colorful edge in (A). Connections among the same type of cells are represented by the same color. (**B**) Average number of cell-cell connection edges in the three groups over time. (**C**) Average number of connection edges among different cell types. (**D**) Average number of connection edges between $GATA3^+$ cells and T^+ cells.

described by analytical equations. To address this, autoencoder-type neural networks (*31*) were used to project embryoid images directly into a low-dimensional latent feature space. Although the latent features extracted by the encoder are less intuitive and interpretable than manually defined morphological and fluorescent features, the latent space is sufficiently comprehensive and informative to allow the decoder to generate realistic synthetic images. As a result, the latent feature embeddings can more accurately reflect the developmental dynamics of the human embryoids.

The architecture of the autoencoder-type networks is shown in Fig. 5A. It consists of a CNN encoder, a CNN decoder, a fully connected neural network (FNN) acting as a discriminator on the latent features, and a CNN-based discriminator applied to the generated images. The FNN and CNN discriminators were included for adversarial training, enabling the model to robustly and accurately capture the hidden feature distribution of the small image dataset (see the "Encoder-decoder model for latent features" section in Materials and Methods) (*31*, *32*).

Each single-channel image is encoded into a latent feature embedding of dimension *n*. Four imaging channels were acquired, corresponding to the fluorescent intensities of 4′,6-diamidino-2-phenylindole (DAPI), GATA3, NANOG, and T markers. Therefore, the overall latent feature space has a dimension of $4n \times N_{\text{img}}$, where N_{img} is the total number of embryoid images (fig. S7). Multiple values of *n* were tested. The algorithm converged and produced realistic images for $n = 5$, 10, 20, and 120 (Fig. 5B). Five autoencoder networks were trained and validated (Fig. 5C), and Fig. 5D shows example images generated using different values of *n*.

As *n* decreased, some information was lost during encoding, resulting in blurrier generated images and deviations in intensity distribution compared to the original images. When $n = 5$, the latent embeddings still captured basic morphological features of the embryoid tissues and central cavities. With $n = 20$, the latent space preserved key fluorescent information of cell lineage markers, including both the magnitude and spatial distribution of signal intensities.

Figures 5 (E and F) compares the morphological and cellular fluorescent features between the original and generated images. Morphological characteristics were accurately reproduced using both 5D and 20D latent features. Fluorescent features in the generated images, such as cell count and marker intensities, were generally consistent with the originals. However, because of image blurriness, the number of distinguishable cells was reduced, and the average marker intensity appeared slightly lower.

A principal components analysis (PCA)–based rotation (*33*) was applied to the *n* latent embeddings, producing *n* orthogonal principal components (PCs) for each single-channel image. The characteristics of these latent PCs for $n = 5$ and $n = 20$ are shown in figs. S8 and S9, respectively. The distributions of PCs were influenced by the initial cell seeding densities across the three groups (fig. S8A). However, the temporal trends of PC1, PC2, and PC3—which together capture the majority of data variation—were largely consistent. As shown in fig. S8 (E and G), PC1 and PC2 are positively correlated with embryoid tissue area and central cavity area. Figure S8 (F and H) indicates that large tissue thickness ratio and roundness tend to correspond to PC values near zero. Unlike manually defined features, these PCs represent complex combinations of multiple physical features and are influenced by the design of the encoder-decoder network and the training process (*31*).

The 20D latent features differed substantially across the four fluorescent channels (DAPI, GATA3, NANOG, and T). As shown in

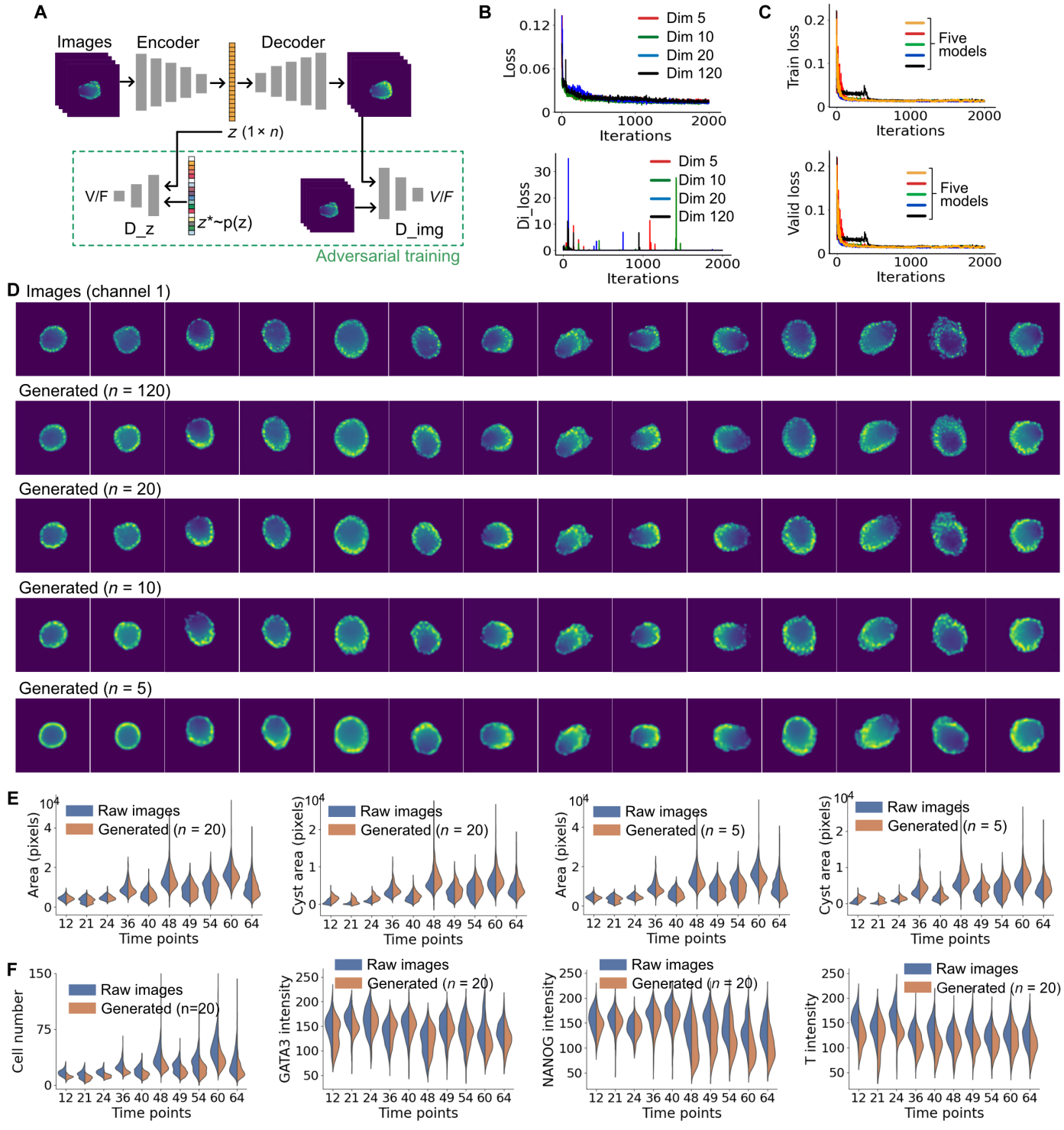


Fig. 5. Latent feature embeddings of embryoid images predicted by the autoencoder neural networks. (**A**) Architecture of the autoencoder neural networks, in which an encoder network extracted low-dimensional latent features from images and a decoder network generated artificial images from the latent embeddings. (**B**) Convergence plots influenced by the latent space dimension *n*. When the latent feature dimension was $n = 5, 10, 20, 120$, the training converged. (**C**) Convergence plots of the training and validation losses, where five neural networks with randomly initialized parameters were trained. (**D**) Comparison of the raw images and generated artificial images influenced by the latent dimension *n*. (**E**) Quantitative comparison of the morphological features (i.e., tissue area and cavity area) between raw images and generated images when $n = 20$ and $n = 5$. (**F**) Quantitative comparison of the cellular features (i.e., number of cells and average cellular intensity) between raw images and generated images when $n = 20$. Note that violin plots at time points 21, 40, 49, 54, and 64 hours are for images in group 1 and group 2, and plots at 12, 24, 36, 48, and 60 hours are for group 3.

fig. S9, PC1 to PC3 were positively correlated with morphological features such as embryoid tissue area, diameter, and major/minor axis lengths. These features were relatively unaffected by the specific fluorescent channel. In contrast, PCs 4, 5, 10, 11, and 19 in the GATA3 and NANOG channels exhibited opposite correlations with many physical features. In addition, PCs 9, 14, and 16 to 19 in the T channel were distinct from those in the other three channels.

In the 5D latent feature space, representative fluorescent images of the embryoids and their corresponding values of PC1 and PC3 are shown in fig. S10. Unlike previous work (*12*), a clear trifurcation of data points into three distinct tissue phenotypes is not observed here. When projecting whole images into the latent space using autoencoder neural networks, the distribution of data points is influenced not only by manually defined phenotypic features but also by subtle or noisy features present in the images. Moreover, the autoencoder model, integrated with adversarial training mechanisms, tends to produce more uniformly distributed latent features, which contributes to the improved quality of the generated images (*31*). Despite this, scatter plots of the latent PCs still offer a preliminary organization of the 3697 images based on similarities in their latent features. For example, in fig. S10, when PC1 and PC3 have large values, the human embryoids tend to be asymmetric, extremely large, and spherical or display outward migration of peripheral cells.

Inference of continuous embryoid development

Each data point in the latent space uniquely corresponds to an embryoid image in the physical space. The development of an embryoid can be visualized as the movement of its corresponding data point along a virtual trajectory in the latent space (Fig. 6A). For multiple embryoid samples, their data points collectively follow characteristic distributions in the latent space. For instance, embryoid images from group 2—acquired at $t = 21$, 40, 49, 54, and 64 hours—can be mathematically represented by data histograms in a 20D orthogonal latent feature space (Fig. 6C). To infer the continuous developmental dynamics of these embryoid samples, a mathematically smooth operator is required to link the temporally separated data histograms.

Here, we propose using a mean-reverting stochastic process based on a mixture of multiple Gaussian distributions as the propagation operator (*34*–*36*). This operator smoothly transforms the data histogram from one time point to another (Fig. 6B). There are many possible choices for the propagation operator, as it does not need to be time-invariant. Its only requirement is to match the data histograms at the five time points, at which human embryoids were collected and examined. We selected the mean-reverting process of mixed Gaussians because it is both simple and widely adopted in generative deep learning models (*36*). Specifically, we assumed the existence of $\mathcal{K}$ distinct embryoid types, with $\mathcal{K}$ remaining constant throughout the 64-hour culture period. In the latent space, the distribution of latent features (PC1 to PC20) for embryoids of type k ($k \in [1, \mathcal{K}]$) at time point t is modeled as a multivariate Gaussian distribution with mean vector $\mu_{k,t}$ and covariance matrix $\sigma_{k,t}$. Accordingly, at each time point t, the real data histograms are approximated as

$$p_t(\boldsymbol{x}) = \prod_{k=1}^{\mathcal{K}} \omega_{k,t} \mathcal{G}\left(\boldsymbol{x} \mid \mu_{k,t}, \sigma_{k,t}\right) \quad (1)$$

where $\mathcal{G}\left(\boldsymbol{x} \mid \mu_{k,t}, \sigma_{k,t}\right)$ is the Gaussian distributions with the mean $\mu_{k,t}$ and covariance matrix $\sigma_{k,t}$, for type k embryoids at time t. $\boldsymbol{x}$ is a 20D vector of latent PC values. $\omega_{k,t}$ is the weight of Gaussian distribution. On the basis of the data histograms in Fig. 6C, $\mu_{k,t}$ and $\sigma_{k,t}$ can be estimated using the expectation-maximization (EM) algorithm at time points $t = 21$, 40, 49, 54, 64 hours (see the "Approximation of data histograms by mixed Gaussian distributions" section in Materials and Methods) (*35*).

To infer the continuous evolution of the distribution function $p_t(\boldsymbol{x})$ of latent PCs, we assumed that latent PC values of type k embryoids $\boldsymbol{x}_k$ followed the mean-reverting stochastic dynamics. Mathematically, the dynamics is described by the stochastic differential equation (SDE) as (*36*)

$$d\boldsymbol{x}_k = \theta_k\left(\mu_{k,t} - \boldsymbol{x}_k\right)dt + \sigma_{k,t}dW \quad (2)$$

where $\mu_{k,t}$ and $\sigma_{k,t}$ are the estimated means and variances at the five discrete time points. θ_k is a model parameter that indicates the ability to be deviated from the mean value $\mu_{k,t}$. $\theta_k = 1$ was used in this study. dt is the time step. $dt = 1$ hour was used in this study. dW is a standard Wiener process. The SDE was numerically solved for the latent PC values in the finer time scales (see the "Mean-reverting stochastic dynamics" section in Materials and Methods).

On the basis of the Bayesian information criterion, two Gaussian components were sufficient to accurately approximate the real data histograms (Fig. 6D). As shown in Fig. 6C, the estimated distributions of 20 latent PCs (red curves) modeled as weighted combinations of two Gaussian components (black curves) closely matched the empirical data histograms (gray areas). These results suggest that the 1424 embryoids in group 2 can be categorized into two distinct types with different characteristics. The estimated parameters of weights, means, and variances for the Gaussian distributions are shown in Fig. 6 (E and F) and fig. S11, respectively. The proportions of the two embryoid types fluctuated ~0.5 (Fig. 6E). Notably, the mean values of PC1 to PC3 exhibited the most significant changes over time (Fig. 6F).

According to the correlation maps in fig. S9, PC1 to PC3 were associated with various morphological features such as embryoid tissue area, diameter, and perimeter, indicating that morphological properties predominantly influenced the evolution of distributions in the latent space. The enlargement of embryoid tissues and central cavities, along with other morphological changes, emerged as key indicators of embryoid development. Both embryoid types, represented by the two Gaussian components, exhibited similar temporal trends in PC1 to PC3. Some PCs—such as PC7, PC9, and PC10—displayed relatively large variations in their mean values for $t < 40$ hours (Fig. 6F). For instance, PC7 showed correlation patterns similar to phenotypic features from the NANOG and DAPI channels but distinct from those of GATA3 and T (fig. S9). Likewise, PC9 was specifically correlated with T-channel features and differed from the other channels. These observations suggest that the expression of NANOG and T underwent changes during early development.

In addition, PCs such as PC14, PC15, and PC18, associated with GATA3 and T (fig. S9), exhibited marked variation around $t =$ 49 hours. Furthermore, the variances of all 20 PCs in Gaussian component 1 increased progressively over time, while in component 2, most variances increased before $t = 54$ hours and then decreased between $t = 54$ to 64 hours (fig. S11).

It should also be noted that the solutions obtained from the EM algorithm are not unique. The algorithm often converged to different local minima, each corresponding to a distinct set of parameters for the two Gaussian components (i.e., $\omega_{k,t}$, $\mu_{k,t}$, and $\sigma_{k,t}$). For instance, fig. S12 presents two different sets of Gaussian components

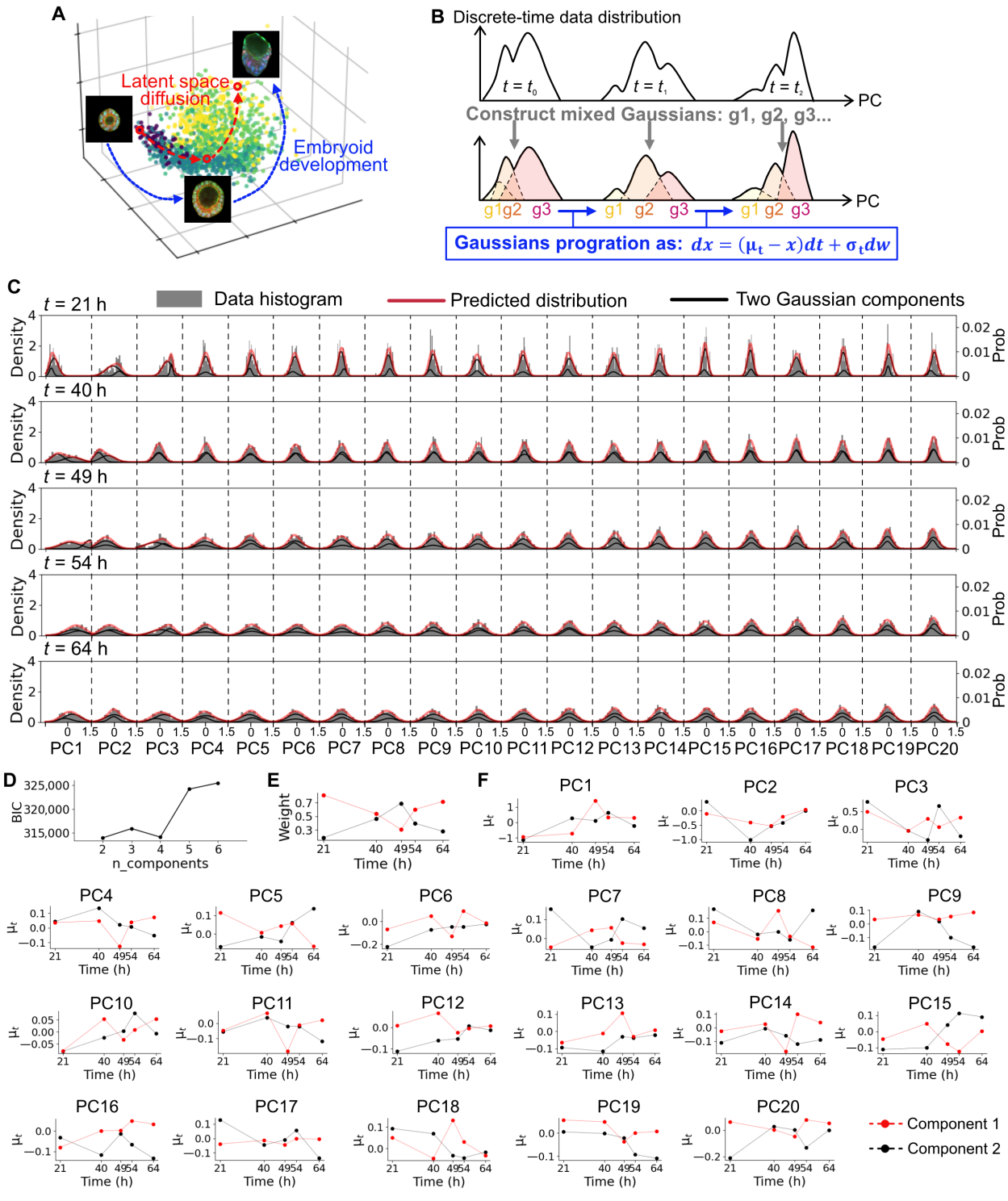


Fig. 6. Approximation of latent feature distributions using mixed Gaussian distributions. (**A**) Schematics showing the relation between evolution of latent feature embeddings and phenotypic development of embryoids. (**B**) Multiple mixed Gaussian distributions for approximating the distributions of features at discrete time points. The mean-reverting stochastic dynamics used to model the continuous propagation of Gaussian distributions. (**C**) Distributions of the 20 latent features (PC1 to PC20) over time. Gray histograms are distributions of latent PCs of the real image data. Black curves are two Gaussian components estimated using the EM algorithm. Red curves are the predicted distribution of latent features based on Eq. 1. (**D**) Bayesian information criterion (BIC) used to select a proper number of Gaussian components. (C) Weights $\omega_{k,t}$ estimated based on the data histograms. (D) Parameters of the two Gaussian distribution components (i.e., means $\mu_{k,t}$) estimated based on the data histograms.

that can both approximate the data histograms effectively. Although the specific values of $\omega_{k,t}$, $\mu_{k,t}$, and $\sigma_{k,t}$ for the 20 PCs varied across these solutions, their overall trends over time remained consistent (see Fig. 6 and figs. S11 and S12). Example raw images of the two types of human embryoids are shown in fig. S13. On average, type 1 embryoids were larger in size compared to type 2, and a greater number of well-developed embryoids were observed in type 1 at $t = 54$ and 64 hours.

Furthermore, continuous developmental trajectories of human embryoids in the latent space were modeled using the SDE described in Eq. 2. Figure 7A shows 10 example trajectories of PC1. For each PC, 10^5 trajectories were generated, and the distributions of PC values at five discrete time points were compared with those of the real data. As illustrated in Fig. 7C, the distributions of PC1 and PC2 derived from the SDE closely matched the empirical data, indicating that the continuous dynamics predicted by the SDE can reliably capture the evolution of latent features in real human embryoids.

By combining the SDE-predicted trajectories with the decoder network, a large number of artificial but realistic embryoid images were generated. Representative examples are shown in Fig. 7B. Key phenotypic features in the generated images—including embryoid tissue area, cell number, and GATA3 intensity—were quantified. Figure 7D presents the distributions of these features at multiple time points. When $t = 21$, 40, 49, 54, and 64 hours, the distributions of the generated images' phenotypic features were found to closely align with those of the real images.

These results suggest that the continuous developmental process of the human embryoid, constructed using the mean-reverting SDE, provides a plausible model for the dynamics underlying human embryoid growth and differentiation. From a biophysical standpoint, the mean-reverting dynamics of mixed Gaussians imply that: (1) multiple distinct developmental paths exist; (2) each path follows a time series of unique, deterministic phenotypic characteristics; and (3) in practice, the development of an embryoid follows a mean-reverting stochastic process along one of these developmental trajectories.

DISCUSSION

In this work, we established an embryoid dataset comprising more than 3000 fluorescent images, along with corresponding segmentation

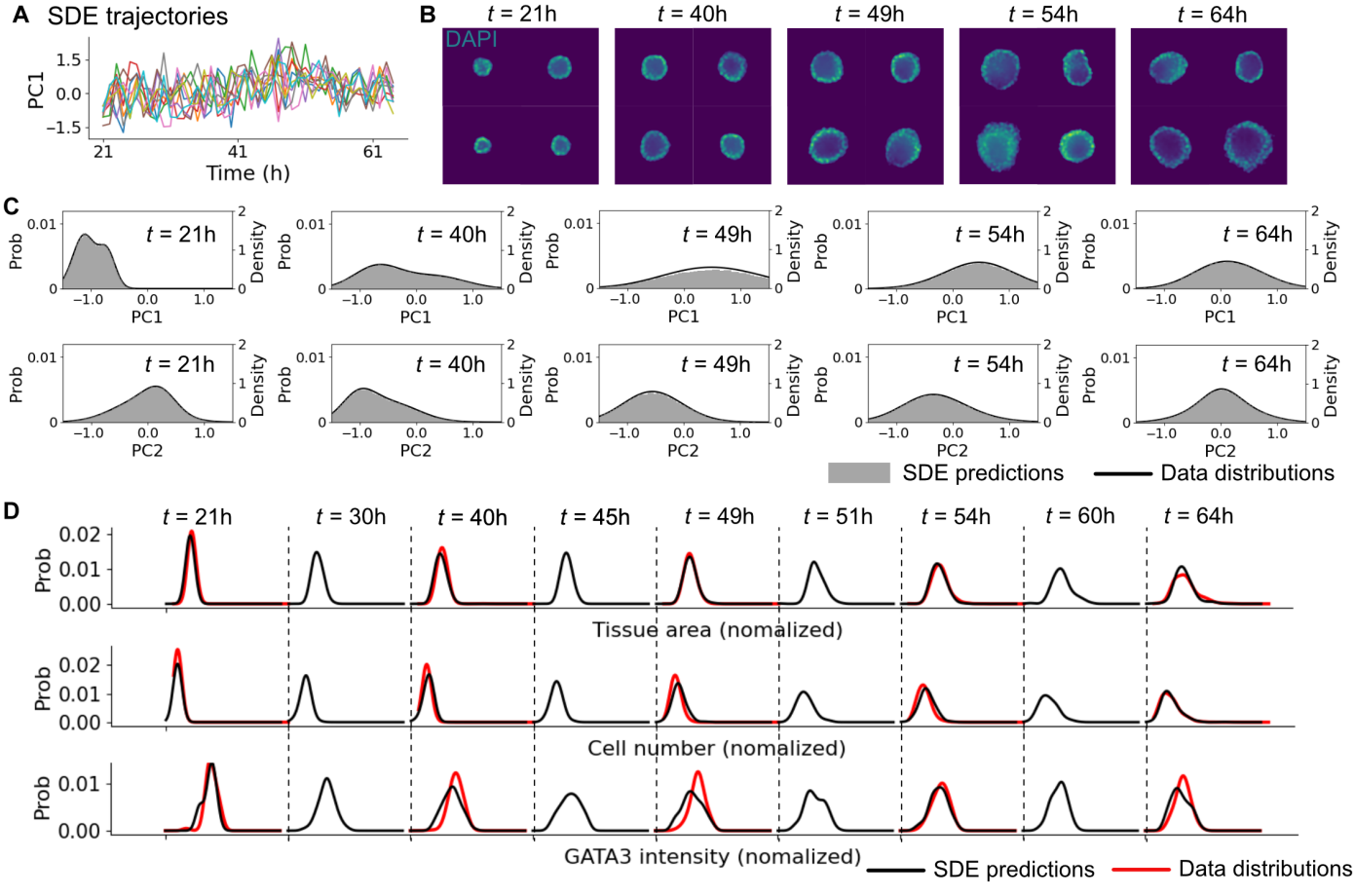


Fig. 7. Inference of the continuous-time embryoid development by constructing an underlying mean-reverting stochastic process to regulate the latent feature evolution. (**A**) Example trajectories of the latent feature PC1 over time generated by solving the mean-reverting stochastic dynamical equation (Eq. 2). (**B**) Examples of the generated images from the decoder neural network based on the continuous-time trajectories of PC1 to PC20 predicted by the SDE. (**C**) Comparison of the PC1 and PC2 distributions, where the black curves are predicted by solving the SDE and the gray histograms are calculated on the basis of the image data. (**D**) Comparison of the morphological and cellular feature distributions (i.e., area, number of cells, and average cellular intensity of GATA3) of the generated artificial images and real images. Red curves are the distribution of real image features only at 21, 40, 49, 54, and 64 hours. Black curves are the feature distributions of generated images at a finer temporal resolution.

masks that identify the shapes of embryoid tissues, cavities, and individual cells. The images were categorized into three groups based on the initial cell seeding densities: small, intermediate, and large. Each embryoid image included four fluorescent channels immunostained with DAPI and antibodies for GATA3, NANOG, and T. Using this dataset, we applied a custom-designed image analysis pipeline to quantify the morphological and cell-level molecular features of the embryoids. Cell-cell connectivity networks were constructed to investigate the spatial distribution patterns of different cell types. We first calibrated the distinct dynamics of tissue growth and lineage diversification in the human embryoids as influenced by the initial cell numbers. Around $t = 40$ to 49 hours of culture, we observed the onset of cell differentiation in the human embryoids, resulting in a "salt-and-pepper" distribution of $GATA3^+$ and T^+ cells. Subsequently, by $t = 48$ to 54 hours, the spatial segregation of $GATA3^+$ and T^+ cells were observed in the human embryoids, with $GATA3^+$ and T^+ cells accumulating on opposite sides of the embryoid tissues, likely involving active cell sorting. By $t = 60$ to 64 hours, the surrounding tissue thickness near $GATA3^+$ cells had become significantly decreased.

Beyond phenotypic analysis, we developed a physics-integrated deep manifold learning framework to capture the latent dynamics of embryoid development based on whole-image information. Latent feature embeddings were extracted directly from the embryoid images using an encoder-decoder neural network. Our results showed that 5D latent features effectively captured morphological characteristics of the human embryoids, while 20D features sufficiently represented both morphological and cellular lineage marker information. Within the 20D latent space, we modeled the developmental progression of the human embryoids as a continuous, stochastic process, a mean-reverting propagation of mixed Gaussian distributions, which captured the underlying regulatory dynamics of human embryoid development. Integrating this model with the neural network allowed us to generate artificial but realistic embryoid images at finer temporal resolutions. The morphological and molecular features of these generated images statistically matched those of the real images of human embryoids collected at discrete time points.

In summary, this work makes two key contributions. First, we present a rich and well-annotated human embryoid dataset, encompassing embryoids cultured from varying initial cell numbers and imaged across multiple time points. Each image includes accurate segmentations of embryoid tissue, fluid-filled cavities, and individual cells, providing a valuable resource for further investigation into multicellular dynamics including morphogenetic changes and cell fate decisions during early human development. Second, we introduce a flexible, physics-integrated deep manifold learning framework capable of modeling complex stochastic developmental dynamics of human embryoids. This whole image–based approach enables the analysis of how underlying regulatory processes drive variations in tissue morphologies and cell lineage diversification patterns, offering insights into the mechanisms guiding the plasticity and robustness of human development. The image processing and analysis pipelines developed here can be adapted and extended to 3D confocal tissue images and time-lapse imaging data from other biological and biomedical studies.

MATERIALS AND METHODS

Generation of hESC-based embryoids

H9 hESC line was maintained in a standard feeder-free culture using mTeSR1 medium (STEMCELL Technologies) and lactate dehydrogenase–elevating virus free hESC-qualified reduced growth factor–based membrane matrix Geltrex (Thermo Fisher Scientific) per manufacturer's instructions. H9 cells were used before P70.

First, polydimethylsiloxane-based pyramidal devices were fabricated by replica molding using AggreWell 400 (STEMCELL Technologies). Next, single-cell suspension of hESCs supplemented with 10 μM Y27632 (Tocris) was added to the pyramidal devises and incubated for 6 hours. Spherical hESC aggregates were then transferred out of the devices and resuspended in the mTeSR1 medium supplemented with 5% Geltrex. Last, these hESCs aggregates were cultured in the low adhesive 96-well plate in a 3D suspension culture manner for 64 hours.

Acquisition of embryoid images

At multiple culture time points (i.e., 12, 21, 24, 36, 40, 48, 49, 54, and 64 hours), a proportion of embryoids were collected from the well plate and stained with fluorescent markers for DAPI (Thermo Fisher Scientific), NANOG (species: rabbit, dilution: 1:200, catalog no. 4903S, vendor: Cell Signaling Technology), GATA3 (species: mouse, dilution: 1:100, catalog no. MA1-028, vendor: Thermo Fisher Scientific) and T (species: goat, dilution: 1:200, catalog no. PA5-46984 or 81694s, vendor: Thermo Fisher Scientific or Cell Signaling Technology). Specifically, embryoid tissues were fixed in 4% paraformaldehyde for 3 hours and permeabilized in 0.1% SDS (sodium dodecylsulfate) for 3 hours. Then, embryoid tissue samples were blocked in 4% donkey serum solution at 4° C for 24 hours before incubation with a primary antibody solution at 4° C for another 24 hours. Samples were labeled with donkey-raised secondary antibodies (1:500 dilution) at 4° C for 24 hours. Four-channel fluorescent images were acquired using an Olympus IX81 fluorescent microscope at × 20 magnification.

Image processing pipeline

An image processing pipeline was developed, which included image rotation and alignment, size adjustment, annotation and segmentation. First, boundaries of embryoid tissues and cavities in the embryoid images (DAPI channel) of group 1 and group 2 were manually annotated using an open-source software VIA (VGG image annotator) (*37*). There are a total of 2219 images of DAPI channel. The pixel-wise areas of singles cells in the embryoid images (DAPI channel) of group 2 (total 1000 images) were manually annotated using an open-source software CVAT (Computer Vision Annotation Tool) (*38*). The annotation results were automatically read using custom python scripts.

Next, a transformer-CNN model developed in the previous work (*30*) was modified and applied for segmentation of embryoid tissues, cavities, and single cells. The model architecture (Fig. 2B) contained ResNet50 for extracting image features, which were inputted into the 12 transformer layers sequentially. The hidden features from both ResNet blocks and the last transformer layer were combined, reshaped, and inputted into the decoder block. The information then went through the CNN layers and the upsampling blocks to generate masks for tissues, cavities, and single cells. The pretrained ResNet50 and transformer layers provided in previous work (*30*) were used. Annotated embryoid images were used to train and validate the transformer-CNN model. The size of the images was adjusted to $4 \times 288 \times 288$ pixels, where each embryoid image contained four fluorescent channels, before inputting into the neural network model.

The predicted masks of tissues and cavities were as accurate as the manual annotations and did not need any modifications. For the

predicted masks of single cells, several morphological modifications were made because the model performance could be influenced by immunostaining artifacts. For example, if the number of pixels in the segmented area was smaller than 20 pixels (i.e., smaller than the average size of cells), then it was regarded as an immunostaining artifact and was removed. If the segmented area contained two or more overlapping cells, then these cells were segmented again on the basis of the curvatures of the boundaries and the Watershed algorithm. A further validation was performed on the basis of embryoid images without annotations. The annotators visualized the raw images and predicted masks image-by-image, ensuring that the predicted masks can represent the boundaries of tissues, cavities, and single cells correctly.

Last, the centers of embryoid tissues were moved to the center of the images for both raw images and masks. The longest axes of the embryoids were aligned vertically, or the tissue areas with the smallest thickness were aligned to the most left side of the images using the custom python scripts.

The morphological features listed in table S1—such as tissue area, cavity area, eccentricity, diameter, perimeter, axis length, and thickness ratio—were calculated on the basis of the masks of tissues and cavities using the Scikit-Image package in python. The cell-wise molecular features listed in table S1—such as cell number; the numbers of $GATA3^+/NANOG^+/T^+$ cells; and the average cellular fluorescent intensity of GATA3, NANOG, and T—were calculated using the masks of single cells. The criteria used for identification of a marker-positive cells were shown in fig. S6 and defined in table S1. The cell-cell connectivity network was constructed on the basis of the information of cell centers and cell types using the self-devised python programs. The violin plots of feature distributions in Figs. 3 to 5 were based on the Seaborn package in python.

Encoder-decoder model for latent features

The autoencoder model proposed in previous work (*31*) was modified and used here, which contained a CNN encoder, a CNN decoder, an FNN discriminator on latent features and a CNN discriminator on generative images. When the latent dimension is $n = 120$, the encoder contained six convolution layers with 64, 128, 256, 512, 1024, and 2048 feature maps, all with 3×3 kernels, rectified linear unit (ReLU) activations, and alternating stride size of 2. The decoder contained seven deconvolution layers with 512, 256, 128, 64, 32, 16, and 1 feature maps. The first six layers used 3×3 kernels, ReLU activations, and stride size of 2. The last layer used a 1×1 kernel, ReLU activations, and stride size of 1. The FNN discriminator contained three FNN layers with 64, 32, and 16 feature maps; batch normalization; and ReLU activation. The CNN discriminator contained four convolution layers with 16, 32, 64, and 128 feature maps, batch normalization and ReLU activations and two FNN layers with 1024 and 1 feature maps, batch normalization and LeakyReLU activations. When the latent dimension is $n = 5, 10, 20$, the encoder contained five convolution layers with 64, 128, 256, 512, and 1024 feature maps, all with 5×5 kernels, ReLU activations, and stride size of 2. The decoder contains seven deconvolution layers with 512, 256, 128, 54, 32, 16, and 1 feature maps using 5×5 kernels for the first six layers and a 1×1 kernel for the last layer. Discriminators on latent features and generated images remained the same.

The minimum-maximum loss function was used for adversarial training, which is expressed as

$$\begin{aligned}\min_{E,G}\max_{D_z,D_{\text{img}}} \quad & \lambda\mathcal{L}\{x, G[E(x), l]\} \\ & + \mathbb{E}_{z^*\sim p(z)}\{\log[D_z(z^*)]\} \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}(\log\{1 - D_z[E(x)]\}) \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}\{\log[D_{\text{img}}(x)]\} \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_{\text{img}}\{G[E(x)]\})]\end{aligned} \tag{3}$$

where $\mathcal{L}(\cdot,\cdot)$ denotes L_2 norm. The encoder E mapped the input image $\boldsymbol{x}$ to a feazture vector, i.e., $E(\boldsymbol{x}) = \boldsymbol{z} \in \mathbb{R}^n$, where n is the dimension of the latent feature space. The decoder (generator) G constructed the output image $\hat{\boldsymbol{x}} = G(\boldsymbol{z}) = G[E(\boldsymbol{x})]$. $z^* \sim p(\boldsymbol{z})$ denotes the random sampling process from a uniform distribution $p(\boldsymbol{z})$ on $[-1, 1]$. $p_{\text{data}}(\boldsymbol{x})$ denotes the distribution of training data. $D_z(\cdot)$ and $D_{\text{img}}(\cdot)$ are outputs from discrimmimnator D_z and image discriminator D_{img}, respectively. λ is the training parameter. Note that a classical autoencoder without adversarial training procedure could also be used to calculate the latent features embeddings. Without adversarial training, only the likelihood function (the first part of Eq. 3) is minimized. However, according to the results in the previous study (*31*), the discriminators of the latent features D_z and generated images D_{img} could help make the training more robust, accelerate the convergence rate, and make the latent feature distributions more uniform. Thus, the adversarial training procedure was used in this study.

In training, the loss function in Eq. 3 was used, where $\lambda = 10^4$, and the four blocks (i.e., encoder, decoder, discriminator of z, and discriminator of generated images) were updated alternatively with a mini-batch size of 64 using the stochastic gradient descent solver adaptive moment estimation ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The learning rate is set as 2×10^{-4}. It took around 2 hours to train the autoencoder model with a NVIDIA RTX A4000 GPU card. Embryoid images of group 1 and 2 are used for training, and images of group 3 are used for validation.

Approximation of data histograms by mixed Gaussian distributions

The latent PC features at time point t corresponding to the embryoid images were calculated using the encoder-docoder network and is denoted as $\boldsymbol{x}_t$. The distribution histograms of $\boldsymbol{x}_t$ was shown in Fig. 6C, which were not Gaussian distributions especially at the early time points and for the first five PCs. To approximate these data distribution histograms, mixed Gaussian model was modified and applied. Specifically, we assumed there were $\mathcal{K}$ types of embryoids in the samples, and $\mathcal{K}$ was a fixed number in the total 64 hours. For each type k ($k \in [1, \mathcal{K}]$), distributions of the latent features of type k embryoids followed multivariate Gaussian distributions, denoted as $\mathcal{G}(\boldsymbol{x} \mid \mu_{k,t}, \sigma_{k,t})$, where $\boldsymbol{x}$ is a vector of latent features, $\mu_{k,t}$ is a vector of means, and $\sigma_{k,t}$ is the covariance matrix of the kth Gaussian distribution. At time t, the distribution of latent PC features is a weighted combination of the $\mathcal{K}$ Gaussian distributions, expressed as $p_t(\boldsymbol{x}) = \sum_{k=1}^{\mathcal{K}} \omega_{k,t}\mathcal{G}(\boldsymbol{x} \mid \mu_{k,t}, \sigma_{k,t})$, where $\omega_{k,t}$ is a vector of weights.

To identify the highly probable number $\mathcal{K}$ of the Gaussian distributions and calculate the Gaussian parameters $\mu_{k,t}, \sigma_{k,t}$ based on the data, the complete data likelihood was used as (*35*)

$$\mathcal{L}\left(\hat{\omega}_{k,t},\hat{\mu}_{k,t},\hat{\sigma}_{k,t},\hat{c}_t\right)=\prod_t\left\{\prod_{i=1}^{n_t}\prod_{k=1}^{\mathcal{K}}\left[\omega_{k,t}\mathcal{G}\left(\boldsymbol{x}_{i,t}\,|\,\mu_{k,t},\sigma_{k,t}\right)\right]^{\mathbb{1}\left(c_{i,t}=k\right)}\right\} \tag{4}$$

where $\mathcal{G}\left(\boldsymbol{x}_{i,t}\,|\,\mu_{k,t},\sigma_{k,t}\right)$ is the probability of a data point $\boldsymbol{x}_{i,t}$ sampled from the Gaussian distribution with parameters $\mu_{k,t},\sigma_{k,t}$. In this study, $\boldsymbol{x}_{i,t}$ is a 20D vector that contains PC values of ith embryoid image ($i\in\left[1,n_t\right]$, where n_t is the number of embryoid images at time t). $\mu_{k,t}$, $\sigma_{k,t}$ are 20D vectors that contain means and variances of the Gaussian distribution of type k embryoids at time t. Note that, since the latent features extracted from the encoder-decoder networks were projected on the orthogonal space using PCA, all the latent PC features are orthogonal. Thus, the covariance matrix of the multivariate Gaussian distributions becomes diagonal matrix. The diagonal elements were stored in the 20D vector $\sigma_{k,t}$. $\omega_{k,t}$ is a vector of weights of the $\mathcal{K}$ Gaussian distributions. $c_{i,t}$ is the type number of the ith embryoid image. $\mathbb{1}\left(c_{i,t}=k\right)$ is an indicator function, which is equal to one if the ith embryoid image belongs to type k, otherwise is zero.

By maximizing the complete data likelihood $\mathcal{L}$, embryoid images were classified into a proper number of types, where distributions of PC values of each embryoid type followed Gaussian distributions. EM algorithm was used to iteratively estimate the parameters $\hat{\omega}_{k,t},\hat{\mu}_{k,t},\hat{\sigma}_{k,t},\hat{c}_t$ at the discrete time points. In practice, for each time point t, the log-likelihood expression (Eq. 5) was maximized by repeating the following E step and M step until algorithm convergence

$$\log\mathcal{L}=\sum_{i=1}^{n_t}\sum_{k=1}^{\mathcal{K}}\sum_{c_i}\mathbb{E}\left[\mathbb{1}\left(c_i=k\right)\right]\left[\log\omega_{k,t}+\log\mathcal{G}\left(\boldsymbol{x}_{i,t}\,|\,\mu_{k,t},\sigma_{k,t}\right)\right] \tag{5}$$

(1) Initialization step:

Five-dimensional vectors whose element values are uniformly distributed random numbers in $[-1.5,1.5]$ were selected as the initial values of μ_k. Note that the maximum and minimum PC values of the embryoid image data were between 1.5 and −1.5. The element values of five-dimensional vectors σ_k were initialized as one, and weights $\omega_{k,t}$ were initialized as $\frac{1}{k}$.

(2) E step:

The log-likelihood at the iteration step s was updated as

$$\left(\log\mathcal{L}\right)^s=\log\sum_{k=1}^{\mathcal{K}}\exp\left[\log\omega_k^s+\log\mathcal{G}_k^s(\boldsymbol{x})\right] \tag{6}$$

where $\mathcal{G}_k^s(\boldsymbol{x})=-\frac{1}{2}\left[\left(\boldsymbol{x}-\mu_k^s\right)^T\left(\sigma_k^s\right)^{-1}\left(\boldsymbol{x}-\mu_k^s\right)+\log2\pi-\log|\,\sigma_k^s\,|^{-1}\right]$.

The classification probability of embryoid i to type k was updated as

$$p^s\left(c_i=k\right)=\frac{\omega_k^s\mathcal{G}_k^s\left(\boldsymbol{x}_i\right)}{\sum_{k=1}^{\mathcal{K}}\omega_k^s\mathcal{G}_k^s\left(\boldsymbol{x}_i\right)} \tag{7}$$

(3) M step:

The weights were updated as

$$\omega_k^{s+1}=\sum_{i=1}^{n_t}\frac{\omega_k^s\mathcal{G}_k^s\left(\boldsymbol{x}_i\right)}{\sum_{k=1}^{\mathcal{K}}\omega_k^s\mathcal{G}_k^s\left(\boldsymbol{x}_i\right)} \tag{8}$$

The means of the Gaussian distributions of type k embryoids were updated as

$$\mu_k^{s+1}=\frac{\sum_{i=1}^{n_t}\omega_k^{s+1}\boldsymbol{x}_i}{\omega_k^{s+1}} \tag{9}$$

The variances were updated as

$$\sigma_k^{s+1}=\frac{\sum_{i=1}^{n_t}p^s\left(c_i=k\right)\left(\boldsymbol{x}_i-\mu_k^{s+1}\right)^2}{\sum_{i=1}^{n_t}p^s\left(c_i=k\right)} \tag{10}$$

Mean-reverting stochastic dynamics

Since the embryoid samples were collected at only a few discrete time points, the mixed Gaussian distributions and their corresponding means and variances $\mu_{k,t}$ and $\sigma_{k,t}$ estimated using the data histograms were not able to reflect the continuous evolution dynamics. A continuous-time stochastic propagation operator needs to be constructed to link these discrete distributions. Note that the propagation of a distribution from time point t_1 to t_2 does not need to be time-invariant, meaning that the distributions at any time point t is not stable but will keep changing over time. Also, changes in the distributions does not need to be time-reversible. Thus, there are many choices of the underlying stochastic propagation processes of the $\mathcal{K}$ Gaussians, such that their means and variances at the five times points (21, 40, 49, 54, and 64 hours) can be matched to the $\mu_{k,t}$ and $\sigma_{k,t}$ estimated from the data. The most straightforward process maybe the advection and diffusion process, where the advection leads to the variation of mean values $\mu_{k,t}$ of the Gaussians, and diffusion coefficient influences $\sigma_{k,t}$. However, because the advection and diffusion during time interval $\left[t_1,t_2\right]$ can influence not only the distribution at time t_2 but also the distributions at the following time points, t_3, t_4, ... it can be computationally hard to determine the proper advection speed and diffusion coefficients, given only the information of $\mu_{k,t}$ and $\sigma_{k,t}$ at five discrete time points. Therefore, in this work, a mean-reverting stochastic process was used due to the simplicity of parameter estimation, and it was widely used in the generative diffusion models (*36*).

The forward equation of the mean-reverting stochastic process was expressed as Eq. 2. The model parameter $\theta_k=1$ was used. A Monte Carlo simulation was used to calculate the forward propagation equation. For example, at $t=21$ hours, two Gaussian distributions $\mathcal{G}_1^{21}$ and $\mathcal{G}_2^{21}$ with means $\mu_{1,21}$, $\mu_{2,21}$ and variances $\sigma_{1,21}$ and $\sigma_{2,21}$ were estimated from the data histograms using the EM algorithm described in the previous section. Then, 10^5 data points were sampled from the two Gaussian distributions as the initial PC values $\boldsymbol{x}_{1,21}$ and $\boldsymbol{x}_{2,21}$. Then, the means $\mu_{1,40},\mu_{2,40}$ and variances $\sigma_{1,40}$ and $\sigma_{2,40}$ of $\mathcal{G}_1^{40}$ and $\mathcal{G}_2^{40}$ were estimated from the data. During the time interval $t\in(21,40)$ hours, Eq. 2 was solved in a discrete time version as

$$\boldsymbol{x}_{k,t+dt}-\boldsymbol{x}_{k,t}=\theta_k\left(\mu_{k,40}-\boldsymbol{x}_{k,t}\right)dt+\sqrt{\sigma_{k,t}dt}\cdot\mathrm{rand}(\mathbf{0}) \tag{11}$$

where $k\in\{1,2\}$ is the type, $\boldsymbol{x}_{k,t}$ and $\boldsymbol{x}_{k,t+dt}$ are vectors of PC values, which have the initial values as $\boldsymbol{x}_{1,21}$ and $\boldsymbol{x}_{2,21}$. The time step $dt=1$ was used, and rand($\mathbf{0}$) is a vector of random numbers of mean zero

and variance 1. Note that if a large θ_k is used, then the means and variances of $\boldsymbol{x}_{k,t+d}$ would quickly changed from $\mu_{k,21}$ and $\sigma_{k,21}$ to $\mu_{k,40}$ and $\sigma_{k,40}$ before the time t reaches 40 hours.

Ethics statement

The human embryoids generated in this work lacked the trophectoderm lineage and therefore did not constitute integrated or complete models of the human embryo. They lacked the potential to develop into a viable organism. All experiments involving hPSCs were conducted with prior approval from the Human Pluripotent Stem Cell Research Oversight (HPSCRO) Committee at the University of Michigan, Ann Arbor, under the HPSCRO application #1049 with the approval date of 09 January 2023. All procedures complied with institutional and national guidelines, including the 2021 International Society for Stem Cell Research guidelines for research involving human embryo models.

Supplementary Materials

This PDF file includes:

Figs. S1 to S13

Table S1

REFERENCES AND NOTES

1. M. N. Shahbazi, Mechanisms of human embryo development: From cell fate to tissue shape and back. *Development* **147**, dev190629 (2020).
2. C. Gerri, S. Menchero, S. K. Mahadevaiah, J. M. A. Turner, K. K. Niakan, Human embryogenesis: A comparative perspective. *Annu. Rev. Cell Dev. Biol.* **36**, 411–440 (2020).
3. R. C. V. Tyser, E. Mahammadov, S. Nakanoh, L. Vallier, A. Scialdone, S. Srinivas, Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
4. T. Bae, L. Tomasini, J. Mariani, B. Zhou, T. Roychowdhury, D. Franjie, M. Pletikos, R. Pattni, B.-J. Chen, E. Venturini, A. Abyzov, F. M. Vaccarino, Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2017).
5. F. Jianping, A. Warmflash, M. P. Lutolf, Stem-cell-based embryo models for fundamental research and translation. *Nat. Mater.* **20**, 132–144 (2021).
6. Y. Zheng, X. Xue, Y. Shao, S. Wang, S. N. Esfahani, Z. Li, J. M. Muncie, J. N. Lakins, V. M. Weaver, D. L. Gumucio, F. Jianping, Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421–425 (2019).
7. Y. Leqian, Y. Wei, J. Duan, D. A. Schmitz, M. Sakurai, L. Wang, K. Wang, S. Zhao, G. C. Hon, W. Jun, Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **591**, 620–626 (2021).
8. Y. Shao, K. Taniguchi, R. F. Townshend, T. Miki, D. L. Gumucio, F. Jianping, A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* **8**, 208 (2017).
9. G. M. Anand, H. C. Megale, S. H. Murphy, T. Weis, Z. Lin, Y. He, X. Wang, J. Liu, S. Ramanathan, Controlling organoid symmetry breaking uncovers an excitable system underlying human axial elongation. *Cell* **186**, 497–512 (2023).
10. H. Nishimura, T. Tanimura, R. Semba, C. Uwabe, Normal development of early human embryos: Observation of 90 specimens at Carnegie stages 7 to 13. *Teratology* **10**, 1–5 (1974).
11. N. Toulany, H. Morales-Navarrete, D. Čapek, J. Grathwohl, M. Ünalan, P. Müller, Uncovering developmental time and tempo using deep learning. *Nat. Methods* **20**, 2000–2010 (2023).
12. K. Chen, Y. Zheng, X. Xue, Y. Liu, A. M. Resto, H. T. Irizarry, F. Jianping, Branching development of early post-implantation human embryonic-like tissues in 3D stem cell culture. *Biomaterials* **275**, 120898 (2021).
13. E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233–1246 (2019).
14. A. Mang, S. Bakas, S. Subramanian, C. Davatzikos, G. Biros, Integrated biophysical modeling and image analysis: Application to neuro-oncology. *Annu. Rev. Biomed. Eng.* **22**, 309–341 (2020).
15. J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington, A. E. Carpenter, Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
16. A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, D. M. Sabatini, CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Nat. Methods* **7**, 1–11 (2006).
17. C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo, A. E. Carpenter, CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* **16**, e2005970 (2018).
18. J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, K. W. Eliceiri, TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
19. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, X. Buote, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, Ilastik: Interactive machine learning for (bio) image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
20. M. M. Mahe, E. Aihara, M. A. Schumacher, Y. Zavros, M. H. Montrose, M. A. Helmrath, T. Sato, N. F. Shroyer, Establishment of gastrointestinal epithelial organoids. *Curr. Protoc. Mouse Biol.* **3**, 217–240 (2013).
21. B. Phipson, P. X. Er, A. N. Combes, T. A. Forbes, S. E. Howden, L. Zappia, H.-J. Yen, K. T. Lawlor, L. J. Hale, J. Sun, E. Wolvetang, M. Takasato, A. Oshlack, M. H. Little, Evaluation of variability in human kidney organoids. *Nat. Methods* **16**, 79–87 (2019).
22. D. Tuveson, H. Clevers, Cancer modeling meets human organoid technology. *Science* **364**, 952–955 (2019).
23. I. Lukonin, D. Serra, L. C. Meylan, K. Volkmann, J. Baaten, R. Zhao, S. Meeusen, K. Colman, F. Maurer, M. B. Stadler, J. Jenkins, P. Liberali, Phenotypic landscape of intestinal organoid regeneration. *Nature* **586**, 275–280 (2020).
24. J. Betge, N. Rindtorff, J. Sauer, B. Rauscher, C. Dingert, H. Gaitantzi, F. Herweck, K. Srour-Mhanna, T. Miersch, E. Valentini, K. E. Boonekamp, V. Hauber, T. Gutting, L. Frank, S. Belle, T. Gaiser, I. Buchholz, R. Jesenofsky, N. Härtel, T. Zhan, B. Fischer, K. Breitkopf-Heinlein, E. Burgermeister, M. P. Ebert, M. Boutros, The drug-induced phenotypic landscape of colorectal cancer organoids. *Nat. Commun.* **13**, 3135 (2022).
25. Y. Lifeng Xiang, Y. Z. Yin, Y. Ma, Y. Li, Z. Zhao, J. Guo, Z. Ai, Y. Niu, K. Duan, J. He, S. Ren, W. Dan, Y. Bai, Z. Shang, X. Dai, W. Ji, T. Li, A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
26. T. Tan, W. Jun, C. Si, S. Dai, Y. Zhang, E. Nianqin Sun, H. S. Zhang, W. Si, P. Yang, H. Wang, Z. Chen, Y. Ran Zhu, R. H.-B. Kang, L. M. Martinez, E. N. Delicado, W. Travis Berggren, M. Schwarz, Z. Ai, T. Li, H. Deng, C. R. Esteban, W. Ji, Y. Niu, J. C. I. Belmonte, Chimeric contribution of human extended pluripotent stem cells to monkey embryos ex vivo. *Cell* **184**, 2020–2032 (2021).
27. J. V. Veenvliet, A. Bolondi, H. Kretzmer, L. Haut, M. Scholze-Wittler, D. Schifferl, F. Koch, L. Guignard, A. S. Kumar, M. Pustet, S. Heimann, R. Buschow, L. Wittler, B. Timmermann, A. Meissner, B. G. Herrmann, Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* **370**, eaba4937 (2020).
28. S. E. Harrison, B. Sozen, N. Christodoulou, C. Kyprianou, M. Zernicka-Goetz, Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science* **356**, eaal1810 (2017).
29. M. Simunovic, J. J. Metzger, F. Etoc, A. Yoney, A. Ruzo, I. Martyn, G. Croft, D. S. You, A. H. Brivanlou, E. D. Siggia, A 3D model of human epiblast reveals BMP4-driven symmetry breaking. *Nat. Cell Biol.* **21**, 900 (2019).
30. J. Chen, L. Yongyi, Y. Qihang, X. Luo, E. Adeli, Y. Wang, L. Le, A. L. Yuille, Y. Zhou, TransUNet: Transformers make strong encoders for medical image segmentation. *ArXiv* **2102**, 04306v (2021).
31. Z. Zhang, Y. Song, H. Qi, Age progression/regression by conditional adversarial autoencoder. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 5810–5818.
32. A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, B. Frey, Adversarial autoencoders. *ArXiv* **1511**, 05644 (2015).
33. S. Wold, Principal component analysis. *Chemom. Intel. Lab. Syst.* **2**, 37–52 (1987).
34. D. Hallac, P. Nystrup, S. Boyd, Greedy Gaussian segmentation of multivariate time series. *Adv. Data Anal. Classi.* **13**, 727–751 (2019).
35. R. Gerlach, C. Carter, R. Kohn, Efficient Bayesian inference for dynamic mixture models. *J. Am. Stat. Assoc.* **95**, 451 (2000).
36. Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, T. B. Schön, Refusion: Enabling large-size realistic image restoration with latent-space diffusion models, in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, 2023), pp. 1680–1691.
37. A. Dutta, A. Zisserman, The VIA annotation software for images, audio and video, in *Proceedings of the 27th ACM International Conference on Multimedia* (ACM, 2019), pp. 2276–2279.
38. M. Aljabri, M. AlAmir, M. AlGhamdi, M. Abdel-Mottaleb, F. Collado-Mesa, Towards a better understanding of annotation tools for medical imaging: A survey. *Multimed. Tools Appl.* **81**, 25877–25911 (2022).

Acknowledgments
Funding: The computation works were conducted on the basis of the support from National Natural Science Foundation of China (grant no. 12472308) and Yunnan Fundamental Research Projects (grant no. 202501 AU070199). The experimental works of acquiring embryoid images were conducted on the basis of the support from the University of Michigan Mcubed Fund, the National Institutes of Health (R21 NS113518, R21 HD105192, and R21 HD109635), and the National Science Foundation (CMMI 1917304, CMMI 2325361, and PFI 2213845). **Author contributions:** K.C.: Writing—original draft, conceptualization, investigation, writing—review and editing, methodology, resources, funding acquisition, data curation, validation, supervision, formal analysis, software, project administration, and visualization. K.-R.Q.: Conceptualization, writing—review, and editing and visualization. J.N.: Conceptualization, writing—review and editing, and supervision. G.G.: Conceptualization and writing—review and editing. C.Y.: Conceptualization and writing—review and editing. J.F.: Conceptualization and writing—review and editing, methodology, funding acquisition, supervision, project administration, and visualization. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The complete dataset of raw fluorescent images of embryoids, acquired in the experiments, are available in Dryad (https://doi.org/10.5061/dryad.9s4mw6ms0). Several test images and masks are provided with the code in the Zenodo repository (https://doi.org/10.5061/dryad.9s4mw6ms0). All other data needed to evaluate the conclusions in this paper are present in the paper and/or the Supplementary Materials. Codes used for image preprocessing and analyzing latent developmental dynamics were developed in Python3, which are available in the Zenodo repository (https://doi.org/10.5061/dryad.9s4mw6ms0).

Submitted 19 July 2024
Accepted 25 June 2025
Published 6 August 2025
10.1126/sciadv.adr8901