

ORIGINAL ARTICLE

Machine learning-assisted imaging analysis of a human epiblast model

Agnes M. Resto Irizarry¹, Sajedah Nasr Esfahani¹, Yi Zheng¹, Robin Zhexuan Yan¹, Patrick Kinnunen², and Jianping Fu^{1,3,4,*}

¹Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA, ²Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, USA, ³Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA, and ⁴Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

*Corresponding author. E-mail: jpfu@umich.edu

Abstract

The human embryo is a complex structure that emerges and develops as a result of cell-level decisions guided by both intrinsic genetic programs and cell–cell interactions. Given limited accessibility and associated ethical constraints of human embryonic tissue samples, researchers have turned to the use of human stem cells to generate embryo models to study specific embryogenic developmental steps. However, to study complex self-organizing developmental events using embryo models, there is a need for computational and imaging tools for detailed characterization of cell-level dynamics at the single cell level. In this work, we obtained live cell imaging data from a human pluripotent stem cell (hPSC)-based epiblast model that can recapitulate the luminal epiblast cyst formation soon after implantation of the human blastocyst. By processing imaging data with a Python pipeline that incorporates both cell tracking and event recognition with the use of a CNN-LSTM machine learning model, we obtained detailed temporal information of changes in cell state and neighborhood during the dynamic growth and morphogenesis of luminal hPSC cysts. The use of this tool combined with reporter lines for cell types of interest will drive future mechanistic studies of hPSC fate specification in embryo models and will advance our understanding of how cell-level decisions lead to global organization and emergent phenomena. **Insight, innovation, integration:** Human pluripotent stem cells (hPSCs) have been successfully used to model and understand cellular events that take place during human embryogenesis. Understanding how cell–cell and cell–environment interactions guide cell actions within a hPSC-based embryo model is a key step in elucidating the mechanisms driving system-level embryonic patterning and growth. In this work, we present a robust video analysis pipeline that incorporates the use of machine learning methods to fully characterize the process of hPSC self-organization into luminal cysts to mimic the luminal epiblast cyst formation soon after implantation of the human blastocyst. This pipeline will be a useful tool for understanding cellular mechanisms underlying key embryogenic events in embryo models.

Key words: human pluripotent stem cells; synthetic embryology; machine learning; image processing

Received April 9, 2021; revised July 2, 2021; accepted July 5, 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.

INTRODUCTION

Human embryo development is a complex process in which cells go through major reorganization and progressive fate specification. Pre-implantation human development leads to the formation of the blastocyst, a hollow sphere of trophoblast cells with an inner cell mass (ICM) composed of both epiblast cells (i.e. embryonic stem cells), which will later form the embryo proper, and hypoblast cells, which will later go on to form the yolk sac. Once the blastocyst begins implantation into the uterine wall, there are a number of developmental events all working in parallel and affecting each other in ways we still don't understand. These processes include the invasion of trophoblast cells into the uterine wall and their differentiation into cytotrophoblast and syncytiotrophoblasts as well as the development of the epiblast into a luminal rosette structure enclosing a central cavity. Soon after, the epiblast cells next to invading trophoblast cells differentiate into the amnion, with the remaining epiblast cells next to the hypoblast remaining pluripotent, leading to the formation of a bipolar epiblast-amnion tissue. While crucial to a successful pregnancy, these developmental events are difficult to study due to both technical limitations and ethical considerations [1, 2]. For years, researchers have tried understanding human development with the use of animal models including mouse and monkey models [3–7]. Recently, there have been increasing efforts toward the development of *in vitro* models of human development with the use of human pluripotent stem cells (hPSCs) including human embryonic stem cells (hESCs) [8–11].

Studies have shown that hPSCs have an intrinsic property to self-organize and differentiate to form complex *in vivo*-like structures. Leveraging this capability, researchers have successfully created a variety of hPSC-based embryo models that recapitulate key steps in early human development [12–15]. A developmental process of particular interest to our group has been the formation of luminal hPSC cysts and the differentiation of hPSCs into amnion cells. Shao *et al.* [13] were the first to show that hPSCs could differentiate into amnion cells. They engineered a 3D biomimetic platform with a soft gel bed made with the basement membrane matrix Geltrex™ and a 3D matrix overlay made with a low concentration of Geltrex™ diluted in culture medium. In this system, hPSC clusters would undergo lumenogenesis and form luminal structures containing a central cavity. Over time, three types of cysts resulted from luminal hPSC clusters: cyst composed of amnion cells, cysts composed of undifferentiated hPSCs, and asymmetric cysts containing amniotic cells at one pole and undifferentiated hPSCs at the opposite pole (Fig. 1). The percentage of each type of cyst was shown to depend heavily on the initial cell plating density. While BMP-SMAD signaling was found to be important for amnion differentiation, the mechanism(s) that led to the initiation of amnion differentiation in the 3D structure has remained elusive.

The ability of hPSCs to self-organize and differentiate into *in vivo*-like structures in *in vitro* settings posits the existence of endogenous developmental programs. Consequently, a crucial characteristic of *in vivo*-relevant, stem cell-based embryo models is their ability to leverage these programs in order to capture the progressive nature of human development. Triggering these developmental programs, however, is not a trivial endeavor; it requires cell culture environments engineered with correct dimensionality as well as correct mechanical and biochemical properties. Having taken the necessary first step of creating a hPSC model that recapitulates a developmental period of interest, the next step becomes the elucidation of the mechanisms at work in the system. *In vivo*, progressive development

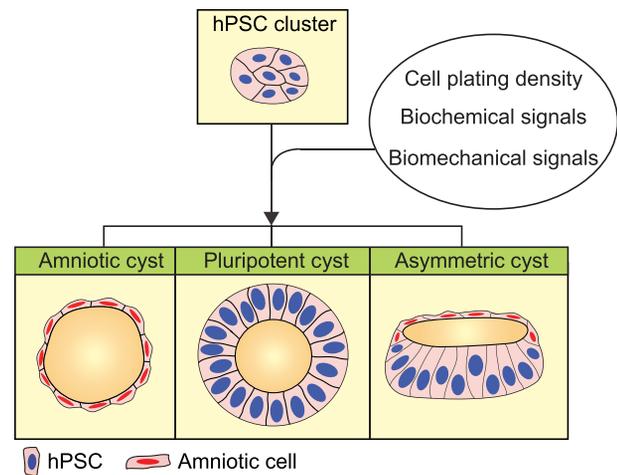


Figure 1. Development of hPSC clusters into three distinct types of luminal structures: amniotic cyst, pluripotent cyst, and asymmetric cyst. Amniotic cysts are composed of amniotic cells, whereas pluripotent cysts are composed of undifferentiated hPSCs. Asymmetric cysts contain amniotic cells at one pole and undifferentiated hPSCs at the opposite pole.

entails branching of distinct lineages and progressive differentiation into cell types with increasingly restricted potential [16]. Studying these processes in hPSC models in a tractable manner requires the use of computational tools that minimize manual curation and bias. Machine learning tools have come to the forefront and are increasingly used to parse the mechanisms at work in these systems. To date, however, many of these efforts have been directed toward the application of single-cell RNA-sequencing (scRNA-seq) data analysis tools [17–21] or examining global features of the observed structures at discrete time points [22, 23]. While these approaches are useful, their discrete nature limits their use for understanding how factors in the local cell microenvironment trigger and guide cell state changes that lead to the emergence of relevant structures. Understanding this requires the ability to continuously monitor individual cells in the system and record division events for later lineage tracing.

There have been several efforts toward the creation of classifiers for the identification of dividing cells. While the methods are varied, the models can be divided into two categories: (i) models that use spatial features [24–26] and (ii) models that use both spatial and temporal features [27–33]. Many of the models that rely only on spatial features for classification utilize morphological feature extraction that leverages the clear differences in visual characteristics between dividing and non-dividing cells [24–26]. Models that lack temporal information, however, face the additional challenge of having to consider how the timing at which the event is captured will affect the features of interest. This is not an issue for spatiotemporal models in which many stages of the division process can be captured and used for the classification. However, rather than focusing on the nucleus, which shows the most obvious visual changes during division, many of the existing models rely on phase contrast microscopy images [27–30, 32, 33] that complicate classification because of the confounding factor of varying cell shape. In this work, we present a computational tool for the comprehensive analysis of live cell imaging data of hPSC cyst formation using a unique nuclear GFP H9 hESC reporter line. Using Python, we created a pipeline that is able to process the images and identify all individual cells in a developing hPSC cyst. The pipeline captures information on both the cell properties and cell neighborhood at

each time point. Further, we trained a machine learning model for event recognition that is able to identify changes in cell state such as division and death by looking at spatiotemporal properties of the nuclei. With this tool, we hope to parse the relationship between the properties of the local environment and cell-level decisions that lead to emergent behaviors like hPSC cyst formation and growth.

MATERIALS AND METHODS

Cell culture substrate preparation

An array of 100 μm -diameter circular adhesive islands was created using a two-step micropatterning method as described previously [34]. Briefly, a poly-dimethylsiloxane (PDMS) elastomeric stamp with an array of circular posts was generated using replica molding from a silicon mold fabricated by standard photolithography and deep reactive ion etching (DRIE) [35, 36]. The center-to-center spacing between adjacent posts on the PDMS stamp was 150 μm , and the post height and diameter were 30 and 100 μm , respectively. The PDMS stamp was coated in 1% Geltrex (Thermo Fisher Scientific; derived from Engelbreth-HolmSwarm tumors similarly as Matrigel[®]) solution for 24 h at 4°C and subsequently rinsed with distilled water and blown dry with nitrogen. Before stamping, the cell culture substrate was prepared by coating a glass coverslip with PDMS and treating it with ultraviolet (UV) ozone (UV-ozone cleaner; Jelight, Irvine, CA) for 7 min to oxidize the PDMS surface. The PDMS stamp was then placed in conformal contact with the PDMS-coated coverslip for 5 s to transfer Geltrex from the stamp to the coverslip. To restrict cell attachment to the circular adhesive islands, the coverslip was treated with Pluronic F127 NF dissolved in PBS (0.2%, w/v ; BASF, Ludwigshafen, Germany) for 1 h at room temperature and rinsed with distilled water. The coverslip was then immersed in mTeSR (STEMCELL Technologies) for a minimum of 2 h to further block the non-functionalized surface of the coverslip. Finally, the coverslip was submerged in mTeSR medium containing 1% Geltrex for 1 h. The coverslip was washed with PBS before cell seeding.

Cell culture

H9 hESCs (WA09, WiCell; NIH registration number: 0062) were used in this study. All culture protocols have been pre-approved by the Human Pluripotent Stem Cell Research Oversight Committee at the University of Michigan. The H9 hESC line is authenticated by the original source, and further authenticated in house by immunostaining for pluripotency markers and differentiation into the three germ layers. Karyotype analysis was performed by Cell Line Genetics. The H9 hESC line was tested negative for mycoplasma contamination (LookOut Mycoplasma PCR Detection Kit, Sigma-Aldrich). H9 cells were cultured in a feeder-free culture system using mTeSR medium and lactate dehydrogenase-elevating virus (LDEV)-free, hESC-qualified reduced growth factor basement membrane matrix Geltrex, per manufacturer's instructions. During each passage, cell culture was visually examined to remove spontaneously differentiated, mesenchymal-like cells. All hESCs used in this work had passage numbers less than P70.

Generation of mTnG cells

For live cell imaging of hESC cyst formation, a membrane tdTomato, nucleus-EGFP (mTnG) H9 hESC line was generated. H2B-EGFP was PCR amplified from a gift plasmid Tcf/Lef:

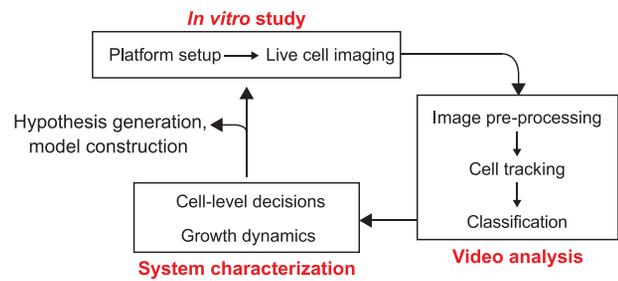


Figure 2. Workflow for live cell imaging data analysis of cyst formation, including *in vitro* experimentation, video analysis, and system characterization.

H2B-GFP (Addgene plasmid #32610). The PCR product was then ligated into the ePiggyBac vector with a constitutively active puromycin selection cassette [37]. membrane-tdTomato was PCR amplified from a gift plasmid pQC membrane TdTomato IX (Addgene plasmid #37351). The PCR product was then ligated into the ePiggyBac vector with a constitutively active neomycin selection cassette [37]. These two plasmids (1.5 μg each) were co-transfected with 1 μg pCAG-PBase (ePiggyBac transposase helper plasmid obtained from Dr Ali H. Brivanlou [37]) using GeneJammer (Agilent Technologies) into H9 hESCs that were plated at 50 000 cells cm^{-2} 24 h prior to transfection. Puromycin selection (2 $\mu\text{g ml}^{-1}$) and G418 selection (250 $\mu\text{g ml}^{-1}$) started at 4 days after transfection. The cells were selected for 7 days. After selection, the cells were dissociated to single cells and replated at low density (400 cells cm^{-2}) for clone picking. A total of 12 clones were hand-picked and evaluated for brightness and pluripotency. Three clones were expanded at the end (mTnG #1, 2, 3). mTnG #1 hESC line has the brightest fluorescent signal and is used in the current study.

Cyst formation assay

Cultured hESC colonies were dissociated into single cells with Accutase (Sigma-Aldrich) at 37°C for 10 min before the cells were centrifuged and re-suspended in mTeSR1 medium containing 10 μM ROCK inhibitor, Y27632 (Tocris), to avoid dissociation-induced apoptosis [38]. Cells were then plated onto coverslips pre-coated with circular adhesive islands at a density of 300 000 cells cm^{-2} . To establish 3D ECM overlay, culture medium was changed to fresh mTeSR1 medium containing 10 μM Y27632 and 4% (v/v) Geltrex 2 h after initial cell seeding. Y27632 was removed 24 h after initial cell seeding, at which time the coverslip was transferred to fluorescence microscopy for live cell imaging.

Live cell video acquisition

mTnG hESCs on the coverslip were imaged using the Zeiss Axio Observer Z1 inverted epifluorescence microscope enclosed in the XL S1 incubator (Carl Zeiss MicroImaging) to maintain cell culture at 37°C and 5% CO_2 . Fluorescence images were recorded with a 20 \times objective for a period of 24 h, with an exposure time of 3 s and a time frame of 10 min to minimize phototoxic effects on cells. A GFP filter set was used for fluorescent imaging of the nuclei of mTnG hESCs.

Image pre-processing

A customized Python program was used to process raw images collected from live cell imaging using fluorescence microscopy. First, contrast was enhanced using adaptive image

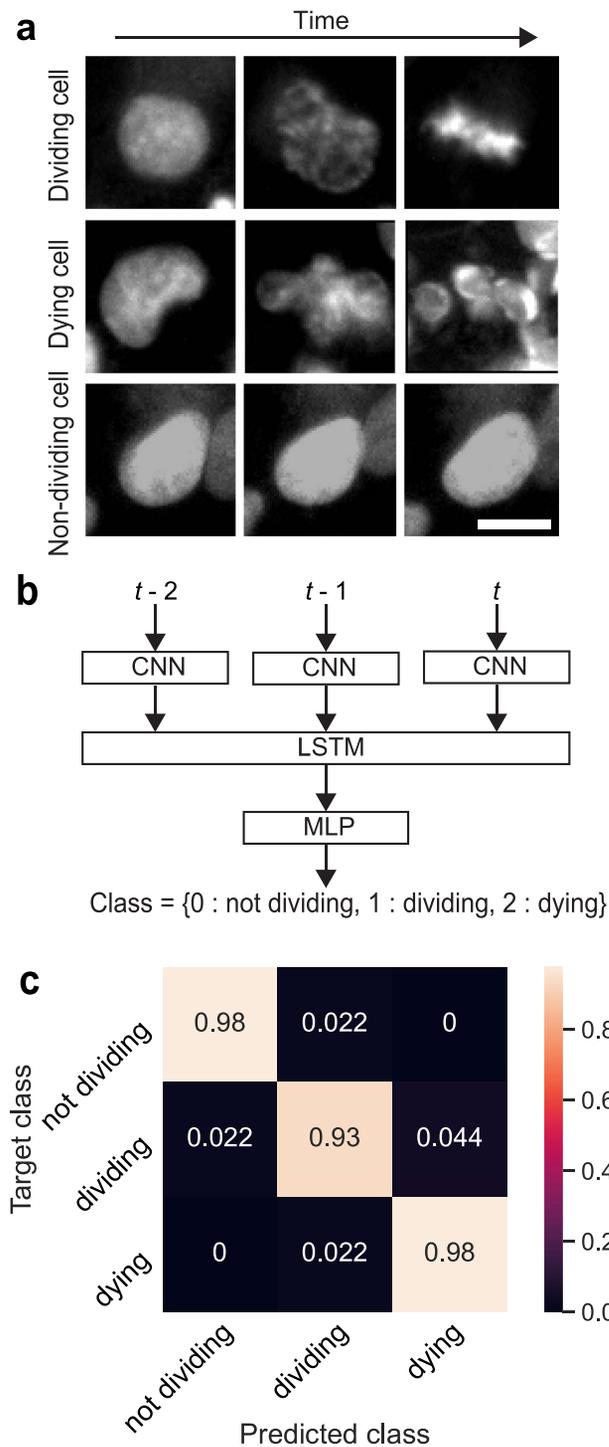


Figure 3. (a) Input image sequences to CNN-LSTM classifier. The top, middle, and bottom rows correspond to a dividing, dying, and non-dividing cell, respectively. Scale bar, 10 μm . (b) CNN-LSTM framework followed by a multilayer perceptron (MLP) for multiclass classification. (c) Confusion matrix for CNN-LSTM classifier.

enhancement developed by Peng *et al.* [39]. Specifically, each pixel in the image is normalized using the mean and variance of a local region surrounding the pixel. This local region is determined adaptively. For a given pixel, the program starts from a given initial size and expands until the standard deviation of the region is equal to or more than a given threshold.

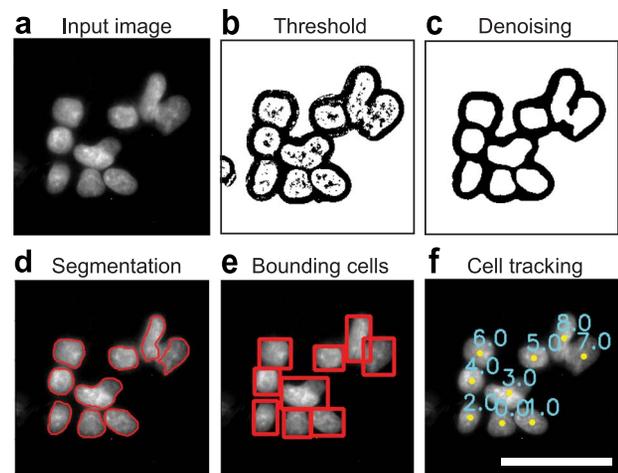


Figure 4. Image processing pipeline. Cell cluster is first isolated from original images (a), before going through adaptive local enhancement and adaptive Gaussian thresholding (b). The image is then denoised with the use of open, erode, and filtering by connectivity (c). Cell clusters are segmented with concavity point analysis and individual cell contours are established (d). Bounding rectangle is then inputted into cell tracker (e), and cells are given a unique ID (f). Scale bar, 50 μm .

For computational tractability, maximum radius was set at 5 pixels. The threshold is in the range of 0.2 to 0.8 and is meant to ensure that the local region has enough relevant structures to classify a pixel as being part of the background or part of an object. After obtaining the region size, the pixel is normalized by subtracting the local mean and dividing by the local standard deviation. This will account for varying background intensity and varying contrast, respectively. A background mask is then obtained by binarizing the resulting image with a binary threshold (`cv2.THRESH_BINARY`). This background mask is further refined with a dilation (`cv2.dilate`, kernel size = [3], iterations = 2) followed by an erosion (`cv2.erode`, kernel size = [5], iterations = 2). Multiplying this resulting background mask by the original image eliminates background noise. Second, contours of cells were identified with the use of adaptive Gaussian threshold (`cv2.adaptiveThreshold` with block size = 23, C = 1) (Fig. 3b). For each pre-processed image, contours were extracted with the use of `cv2.findContours` with `cv2.RETR_TREE` and `cv2.CHAIN_APPROX_NONE`. The third step is to carry out segmentation to find individual cell contours (Fig. 3d). The pipeline measures the area and circularity of each contour. Contours identified as individual cells are stored. The contours identified as cell clusters undergo concavity point-pair segmentation, a method developed by Farhan *et al.* [40] based on finding concavity point-pairs using a variable-size rectangular window. In brief, using an established interval, a list of contour coordinates is first extracted from the binary image of the cell cluster. For each coordinate in the list, lines are drawn to the next two points in the list. Once a line passes through the image background (i.e. a pixel with value 0), the algorithm finds the contour coordinate at which the line no longer passes through the background and establishes this coordinate as a concavity point. After filtering the resulting point list to account for contour irregularities, the program finds the directionality vector of each concave area. Using this vector, each concavity point establishes a rectangular window in which to search for other concavity points. Once all concavity points have paired up, a line is drawn between them and the

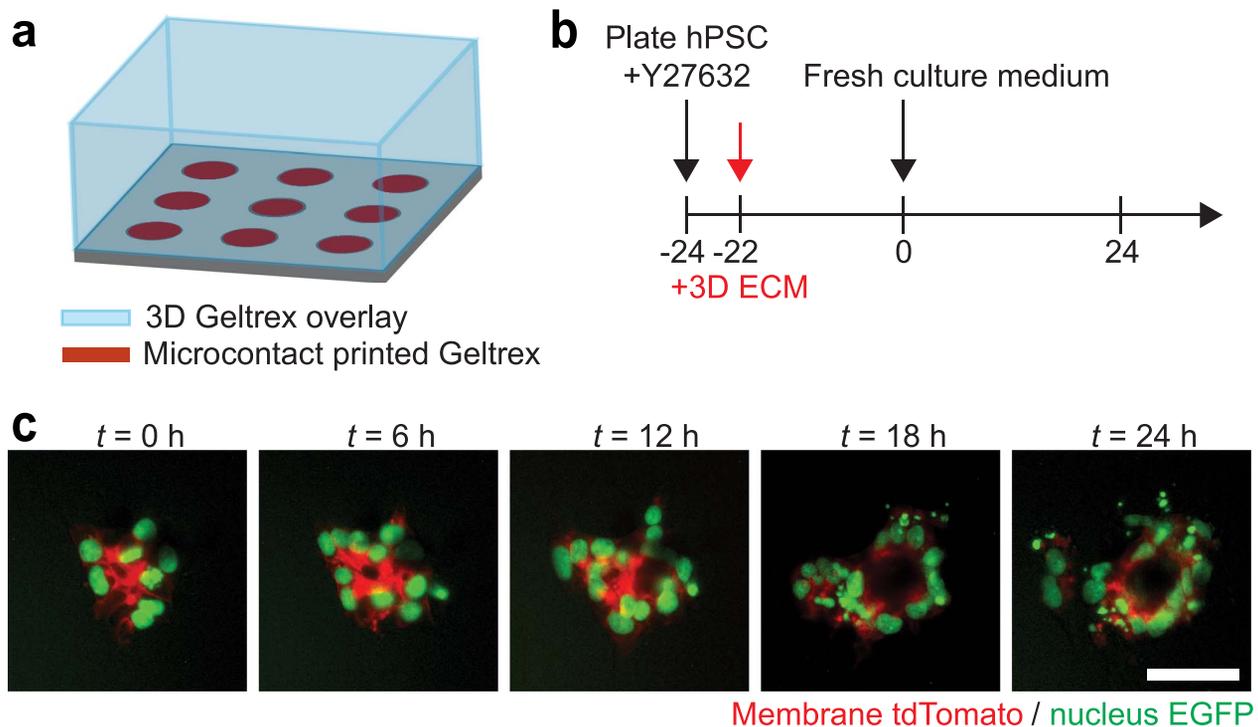


Figure 5. hPSC cyst formation experiment. (a) Cartoon of experimental platform consisting of micropatterned adhesive islands and a 3D Geltrex overlay. (b) Experimental protocol timeline. (c) Live imaging of a hPSC cluster at different time points. In this assay, the mTnG H9 hESC line was used.

cluster is segmented. Farhan *et al.* [40] validated the method with the use of three data sets, two of which contained bright field microscopy images of yeast cells, and one which contained fluorescent microscopy images of yeast cells. They showed that the concavity point-pairing segmentation method was highly effective, with precision averaging at 0.98 [40].

Image selection for CNN-LSTM

The machine learning classifier used was a deep learning model consisting of a convolutional neural network (CNN) connected to a long short-term memory (LSTM) network. The data set for supervised training contained sequences of three time points showing three different classes of cells: dividing, dying, and non-dividing. These sequences of dividing, dying, and non-dividing nuclei were manually cropped from live cell videos (Fig. 3a). The sample set contained 450 samples, with an equal amount of every class.

Parameters for CNN-LSTM

The CNN-LSTM model was constructed using keras.Sequential, which yields a linear stack of layers. The CNN layers consisted of a repeating pattern of convolution, max pooling, and batch normalization followed by one dropout and one global max pooling layer. The CNN model output for each sequence of images was passed on to an LSTM layer via a TimeDistributed layer. This layer extracts features from each image in the sequence and passes it to the LSTM. The final layers in the model create a fully connected network with the use of dense layers. Rectified linear units (ReLU) were used as the activation function in all of the convolutional layers and dense layers, except for the last one. The last dense layer used softmax activation in order to carry out multiclass classification. The model was compiled

using Adam as the optimizer, categorical crossentropy for the loss calculation, and accuracy as the metric evaluated by the model. The number of epochs was 100. During training, callback with ModelCheckpoint was used to store the best model based on validation accuracy. A 70-20-10 split was used to create the training, testing, and validation data sets. Because the data set is small, data augmentation was carried out using a data generator class. Using this generator, the data set was randomly shuffled and images were transformed using rotation (range of 5), shifts in height (range of 0.1), and shifts in width (range of 0.1). Transformations in this data generator class were carried out using the Keras ImageDataGenerator class. One-hot encoding was applied to the labels before training.

Video analysis pipeline

Live cell videos were analyzed with a Python pipeline. Images were pre-processed and all individual nuclei were identified. A cell tracker python class was used to give each cell a unique identification (ID) number and track cells from one time point to another using Euclidean distance. The nuclei were cropped from the image and stored in a Python dictionary. For event recognition, the cropped nuclei of the current time point and the cropped nuclei from the previous two time points were passed as input to the CNN-LSTM classifier. Whenever a new cell would appear in the environment, the parent cell would be identified using a parent score (ps) parameter. Newly divided cells tend to be small, bright, and similar in size. For this reason, the ps takes into account both the classification of the cells in the previous time point and the similarity in area and brightness of the nuclei between the new cell and the possible sister cell. The cell with the highest ps in the local neighborhood of the newly appeared cell would be assigned as the parent. At this point, the daughter cell with the parent ID receives a new ID, and the parent ID of

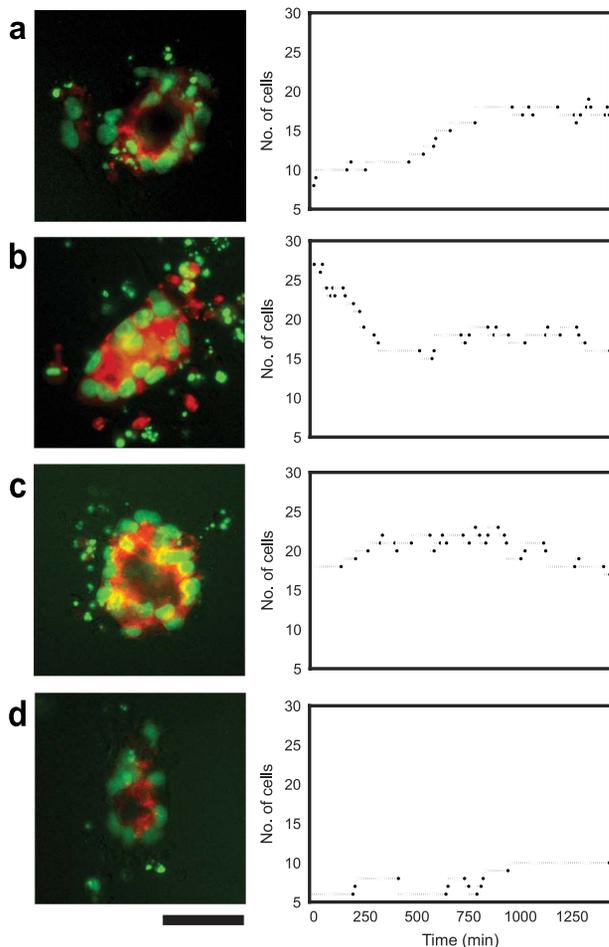


Figure 6. Growth dynamics of four different hPSC cysts. The left and right columns show the cysts after 24 h of growth and the number of cells in the cyst through the period of 24 h, respectively. In this assay, the mTnG H9 hESC line was used.

the daughter cells is stored. The number of neighboring cells and the average distance to neighbors are stored for each cell at every time point. The output of the video cell analysis is a Python DataFrame with the cell IDs, cell positions, number of neighbors, average distance to neighbors and parent IDs for each time point.

RESULTS

Cyst formation analysis pipeline

Analysis of the morphogenesis of a multicellular structure at discrete time points can give insight into the system-level dynamics governing its growth and development. However, important cell-level dynamics and the degree of stochasticity and heterogeneity in a multicellular system remain difficult to elucidate. To better study the dynamic process of hPSC cyst formation, we developed an experimental platform to provide a biomimetic niche for the formation of hESC cysts in a controllable and robust manner. Specifically, an array of 100 μm -diameter circular adhesive islands was created on a coverslip, before hESCs expressing membrane tdTomato and nucleus-EGFP (mTnG) were seeded onto the coverslip. Two hours after cell seeding, culture medium was changed to fresh mTeSR1 medium containing 10 μM Y27632 and 4% (*v/v*) Geltrex, to establish a 3D ECM overlay. The coverslip

was transferred to a Zeiss Axio Observer Z1 inverted epifluorescence microscope 24 h after cell seeding. To track the dynamics of hESC cyst formation, live cell imaging was conducted for 24 h with a depth of focus that captured all the cells in the system. Given their self-organizing property, hESCs confined on adhesive islands on the coverslip formed small clusters and underwent epithelization and lumenogenesis. Throughout cyst formation, cells showed limited movement ability in the z-direction. Live cell imaging data were then processed with the use of a Python pipeline capable of image processing, cell tracking, and event recognition. With these extracted data, a comprehensive characterization of cell states and actions during hESC cyst formation could be conducted, using a workflow that includes image processing with machine learning, and the characterization of cell state and cyst growth (Fig. 2).

Event recognition and image processing

Parsing the relationship between cell actions and their local microenvironment is a necessary step for elucidating the mechanisms that drive hESC cyst formation and development. As a first step to carrying out this analysis, we sought to develop a machine learning model capable of detecting two important changes in cell state: division and death. While CNNs are often used for image classification, we sought to add robustness to the model by also leveraging temporal information with the use of an LSTM network. CNN-LSTM has been utilized for imaging analysis to detect mitotic cells recorded using time-lapse phase-contrast microscopy [30]. Following this logic and utilizing mTnG hESCs that show significant changes in nuclear shape and area for both dividing and dying cells, we trained a CNN-LSTM classifier (Fig. 3b). The classifier identifies three cell states: dividing, dying, and non-dividing (Fig. 3a). While the use of a single image could lead to correct classification, the CNN-LSTM model is able to leverage information on the temporal changes in nuclear shape using live cell imaging data. By using a set of 450 manually labeled images with an equal amount of each class, with a 70-20-10 test-train-validation split and data augmentation, a 96.3% overall accuracy in event recognition was achieved using the CNN-LSTM classifier (Fig. 3c).

Having successfully trained a machine learning model for event recognition, we next sought to create an image processing pipeline capable of identifying individual cell nuclei. While there are a number of computational tools available for identification and tracking of cells in a multicellular system [41–43], identifying and tracking cells within a forming cyst presents unique challenges that require the use of more catered approaches. For example, in our test of the commonly used watershed method for segmentation, it is difficult to carry out correct segmentation of cell clusters, likely because of the compact arrangement of cells in hESC cysts. Additionally, global thresholding methods were unsatisfactory for finding hESC clusters, likely because of variations in nuclear GFP intensity. To address these challenges, we devised an imaging processing pipeline uniquely suited to carry out thresholding and segmentation in tightly packed hESC cysts. The image processing steps consist of thresholding, denoising, segmentation, and identification of individual cells in hESC cysts (Fig. 4). For the segmentation step, we utilized adaptive local enhancement [39] to enhance the contrast between nuclei and background. For binarization, we utilized adaptive Gaussian thresholding to ensure that cells that are slightly out of focus can still be identified. After binarization and denoising, hESC clusters were segmented with the use of concavity point pairing analysis [40]. Cells are given a unique

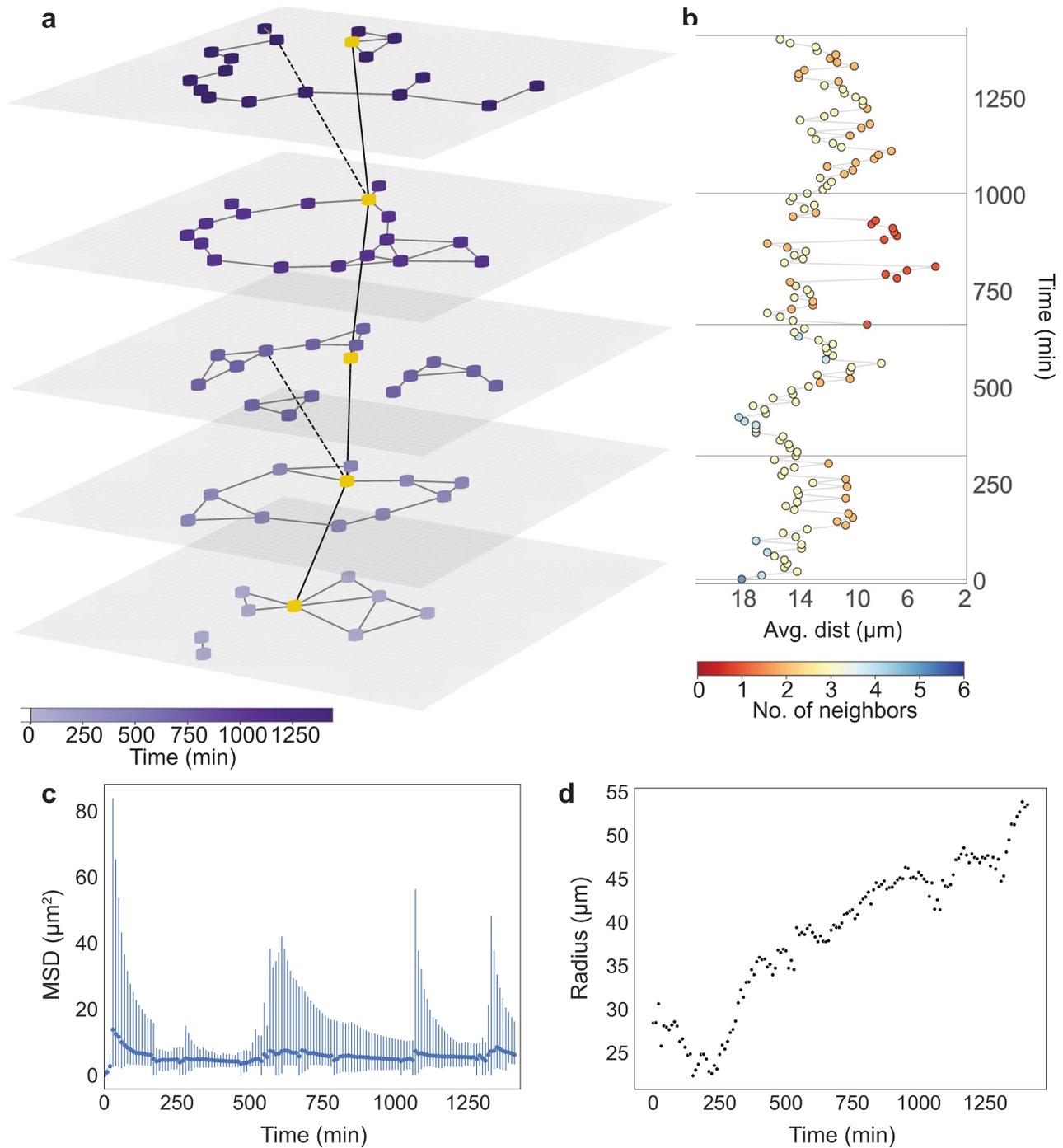


Figure 7. Characterization of hPSC cyst formation dynamics. (a) Network representation of a cyst at different time points. The color of the nodes in each plane indicates the time point. Each node represents a single cell in the cell cluster. Connections between cells are established based on a threshold distance. Cell lineage of the gold-colored cell is shown with the black line connecting the cell at different time points. The dotted lines indicate instances of cell division. (b) Number of neighbors corresponding to the gold-colored cell and average distance between the gold-colored cell and its neighbors as a function of time. (c) Average mean squared displacement (MSD) of all the cells in the cyst as a function of time. The dark blue points and bars represent the average MSD and the range of MSDs, respectively. (d) Radius of the cyst as a function of time.

identification (ID) number and tracked from one time point to another with the use of Euclidean distance.

Live cell data processing and system characterization

As mentioned earlier, the experimental platform for the formation of hESC cysts consisted of an array of micropatterned

Geltrex islands with a diameter of 100 μm generated with a two-step micropatterning process [34] (Fig. 5a). After a period of 24 h in which mTnG hESCs were allowed to attach and cluster onto the adhesive islands, the experimental platform was transferred to a fluorescent microscope for live cell imaging for a period of 24 h (Fig. 5b and c). To avoid cytotoxic effects, images were recorded at intervals of 10 min.

Having obtained live cell data from various hESC cysts, we processed the images using the Python pipeline. Figure 6 shows the growth profiles of four different hESC cysts. The growth profiles vary greatly between the cysts. While there are periods of a sustained increase in cell number like the one seen in Fig. 6a between 500 and 750 min, we can also find periods of a sustained decrease in the number of cells as seen in Fig. 6c between 0 and 250 min. Regardless of the growth profile, however, the number of cells seems to plateau for all the cysts. While the final cell numbers might be similar among the four hESC cysts, there are a number of different growth trajectories that could not have been inferred from looking at the final cyst configuration. In the context of modeling human development, this information facilitates the study of how these changes in growth dynamics correspond to relevant cell specification events. Further, as cells progressively differentiate and more populations appear in a system, we can start to study the growth dynamics of specific populations and how they relate to correct form and function in the structure.

Lineage tracing is a powerful tool for parsing the mechanisms guiding morphogenetic events in a multicellular system. It has many uses including providing insight into the timing of differentiation of cell types of interest and helping identify lineage-specific precursor cells. Combined with the ability to record properties of the local cell microenvironment, it can help parse when and why different cell types arise. Figure 7a shows a network representation of the cyst shown in Fig. 6a at different time points. Edges between cells of a given cyst are connected to each other based on an established threshold distance. The solid line going from one time point to another indicates a chosen cell lineage, with an additional dotted line indicating cell divisions. Having tracked a cell and established its lineage, we can characterize the local cell density experienced by the cells in the lineage throughout time. As can be seen in Fig. 7b, the number of cells in the neighborhood and the average distance from neighbors of the selected cell continue to vary even after cell number in the cyst has plateaued (Fig. 6a). We can also see from Fig. 7d that the cyst radius continues to increase after cell number has plateaued, suggesting that structure growth does not necessarily correlate with increased cell number. From the MSD plot in Fig. 7d we can see that the average MSD in the system remains low throughout the 24 h. This is likely a result of the confinement provided by the adhesive islands in which the cells exist. As different cell types start to arise, this pipeline output can be used to assess differences in movement dynamics between different cell populations and their spatial segregations. For example, mesoderm cells, which are studied in our post-implantation amniotic sac embryoid (PASE) model [13], are known to be more migratory as compared to other populations such as epithelial ectoderm and endoderm cells. While the model presented here is limited to the first 24 h of cyst formation, future efforts can be devoted to extending this timeframe to include important events such as symmetry breaking caused by the appearance of amnion-like cells (AMLCs) in the PASE. In our PASE model, we found that the initial cell seeding density has a significant effect on morphogenesis and differentiation of hESCs [13, 34]. With the information that can be obtained from this pipeline, we can begin to understand the role of initial conditions like cell seeding density, and we can begin the work of relating changes in the local environment with cell-level decisions that lead to the cyst-level growth and patterning.

CONCLUSIONS

The successful generation of human embryo-like structures is a crucial step in advancing fundamental understanding of human development, without using intact, natural human embryos. However, limitations on the insights gained through analysis of human embryo-like structures at discrete time points drove us to create a live cell video processing pipeline catered for the unique challenges of our system. With the use of both spatial and temporal information, we were able to create a machine learning model for event recognition. Furthermore, this model was integrated into an image processing pipeline that leveraged specialized image processing tools for the identification and tracking of individual cells in our system. With this integrative pipeline we were able to characterize the cell states and actions during the dynamic growth and morphogenesis of luminal hESC cysts. Combining this tool with reporter lines for cell types of interest, we hope to advance in our goal to elucidate the mechanisms driving lumenogenesis, cyst growth, and cell fate specification in our *in vitro* hESC models of human development.

FUNDING

A.M.R.I. is partially supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE 1256260. This research is supported by the Michigan-Cambridge Collaboration Initiative, the University of Michigan Mcubed Fund, the 21st Century Jobs Trust Fund received through the Michigan Strategic Fund from the State of Michigan (Grant CASE-315037), the National Institutes of Health (R21 NS113518 and R21 HD100931), and the National Science Foundation (CMMI 1917304 and CBET 1901718).

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Pera MF. Human embryo research and the 14-day rule. *Dev.* 2017;144:1923–5.
2. Rivron N, Pera M. Debate ethics of embryo models from stem cells. *Nature* 2018;564:183–5.
3. Ma H, Zhai J, Wan H et al. *In vitro* culture of cynomolgus monkey embryos beyond early gastrulation. *Science* (80-) 2019;366:eaax7890.
4. Niu Y, Sun N, Li C et al. Dissecting primate early post-implantation development using long-term *in vitro* embryo culture. *Science* (80-) 2019;366:eaaw5754.
5. Sasaki K, Nakamura T, Okamoto I et al. The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev. Cell* 2016;39:169–85.
6. Sozen B, Amadei G, Cox A et al. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat. Cell Biol.* 2018;20:979–89.
7. Sozen B, Cox AL, de Jonghe J et al. Self-Organization of Mouse Stem Cells into an Extended Potential Blastoid. *Dev. Cell* 2019;51:698–712.e8.
8. Resto Irizarry AM, Nasr Esfahani S, Fu J. Bioengineered pluripotent stem cell models: new approaches to explore early human embryo development. *Curr. Opin. Biotechnol.* 2020;66:52–8.

9. Taniguchi K, Heemskerk I, Gumucio DL. Opening the black box: Stem cell-based modeling of human post-implantation development. *J. Cell Biol.* 2019;**218**:410–21.
10. Shahbazi MN, Siggia ED, Zernicka-Goetz M. Self-organization of stem cells into embryos: A window on early mammalian development. *Science (80-.)* 2019;**364**:948–51.
11. Fu J, Warmflash A, Lutolf MP. Stem-cell-based embryo models for fundamental research and translation. *Nat. Mater.* 2021;**20**:132–44.
12. Warmflash A, Sorre B, Etoc F et al. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* 2014;**11**:847–54.
13. Shao Y, Taniguchi K, Townshend RF et al. A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* 2017;**8**:1–15.
14. Simunovic M, Metzger JJ, Etoc F et al. A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. *Nat. Cell Biol.* 2019;**21**:900–10.
15. Zheng Y, Xue X, Shao Y et al. Controlled modelling of human epiblast and amnion development using stem cells. *Nature* 2019;**573**:421–5.
16. Blanpain C, Simons BD. Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* 2013;**14**:489–502.
17. Chen L, Pan XY, Guo W et al. Investigating the gene expression profiles of cells in seven embryonic stages with machine learning algorithms. *Genomics* 2020;**112**:2524–34.
18. Shah N, Li J, Li F et al. An Experiment on Ab Initio Discovery of Biological Knowledge from scRNA-Seq Data Using Machine Learning. *Patterns* 2020;**1**:100071.
19. Peng L, Tian X, Tian G et al. Single-cell RNA-seq clustering: datasets, models, and algorithms. *RNA Biol.* 2020;**17**:765–83.
20. Andrews TS, Hemberg M. Identifying cell populations with scRNASeq. *Mol. Aspects Med.* 2018;**59**:114–22.
21. Liu J, Fan Z, Zhao W et al. Machine Intelligence in Single-Cell Data Analysis: Advances and New Challenges. *Front. Genet.* 2021;**12**:655536.
22. Guo J, Wang P, Sozen B et al. Machine learning-assisted high-content analysis of pluripotent stem cell-derived embryos in vitro. *Stem Cell Reports* 2021;**16**:1331–46.
23. Chen K, Zheng Y, Xue X et al. Branching development of early post-implantation human embryonic-like tissues in 3D stem cell culture. *Biomaterials* 2021;**275**:120898.
24. Irshad H, Jalali S, Roux L et al. Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach. *J. Pathol. Inform.* 2013;**4**:12.
25. Liu A, Hao T, Gao Z et al. Nonnegative mixed-norm convex optimization for mitotic cell detection in phase contrast microscopy. *Comput. Math. Methods Med.* 2013;**2013**:176272.
26. Huang CH, Lee HK. Automated mitosis detection based on eXclusive Independent Component Analysis. *Proc. - Int. Conf. Pattern Recognit.* 2012;**ICPR2012**:1856–9.
27. Yang F, Mackey MA, Ianzini F et al. Cell segmentation, tracking, and mitosis detection using temporal context. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 2005;**3749 LNCS**:302–9.
28. Gallardo GM, Yang F, Ianzini F et al. Mitotic cell recognition with hidden Markov models. *Med. Imaging 2004 Vis. Image-Guided Procecd. Disp* 2004;**661**:5367.
29. Liu AA, Li K, Kanade T. Mitosis sequence detection using hidden conditional random fields. 2010 7th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2010 - Proc 2010;580–3. [10.1109/ISBI.2010.5490279](https://doi.org/10.1109/ISBI.2010.5490279).
30. Su YT, Lu Y, Chen M et al. Spatiotemporal Joint Mitosis Detection Using CNN-LSTM Network in Time-Lapse Phase Contrast Microscopy Images. *IEEE Access* 2017;**5**:18033–41.
31. Harder N, Mora-Bermudez F, Godinez WJ et al. Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time. *Genome Res.* 2009;**19**:2113–24.
32. Nie WZ, Li WH, Liu AA et al. 3D Convolutional Networks-Based Mitotic Event Detection in Time-Lapse Phase Contrast Microscopy Image Sequences of Stem Cell Populations. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2016;52–62. [10.1109/CVPRW.2016.171](https://doi.org/10.1109/CVPRW.2016.171).
33. Li K, Miller ED, Chen M et al. Computer vision tracking of stemness. 2008 5th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, Proceedings, ISBI 2008;847–50. [10.1109/ISBI.2008.4541129](https://doi.org/10.1109/ISBI.2008.4541129).
34. Nasr Esfahani S, Shao Y, Resto Irizarry AM et al. Microengineered human amniotic ectoderm tissue array for high-content developmental phenotyping. *Biomaterials* 2019;**216**:119244.
35. Weng S, Shao Y, Chen W et al. Mechanosensitive subcellular rheostasis drives emergent single-cell mechanical homeostasis. *Nat. Mater.* 2016;**15**:961–7.
36. Fu J, Wang YK, Yang MT et al. Mechanical regulation of cell function with geometrically modulated elastomeric substrates. *Nat. Methods* 2010;**7**:733–6.
37. Lacoste A, Berenshteyn F, Brivanlou AH. An Efficient and Reversible Transposable System for Gene Delivery and Lineage-Specific Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell* 2009;**5**:332–42.
38. Watanabe K, Ueno M, Kamiya D et al. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* 2007;**25**:681–6.
39. Peng JY, Hsu CN, Lin CC. Adaptive image enhancement for fluorescence microscopy. *Proc. - Int. Conf. Technol. Appl. Artif. Intell. TAAI* 2010;**2010**:9–16. [10.1109/TAAI.2010.13](https://doi.org/10.1109/TAAI.2010.13).
40. Farhan M, Yli-Harja O, Niemistö A. A novel method for splitting clumps of convex objects incorporating image intensity and using rectangular window-based concavity point-pair search. *Pattern Recognit.* 2013;**46**:741–51.
41. Schindelin J, Arganda-Carreras I, Frise E et al. Fiji - an Open platform for biological image analysis. *Nat. Methods* 2009;**9**:676–82.
42. McQuin C, Goodman A, Chernyshev V et al. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 2018;**16**:1–17.
43. de Chaumont F, Dallongeville S, Olivo-Marin J, ICY: A new open-source community image processing software. 2011 8th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, Proceedings, ISBI 2011;234–7. [10.1109/ISBI.2011.5872395](https://doi.org/10.1109/ISBI.2011.5872395).