Machine learning-assisted imaging analysis of a human epiblast model

Agnes M. Resto Irizarry[1], Sajedeh Nasr Esfahani[1], Yi Zheng[1], Robin Zhexuan Yan[1], Patrick Kinnunen[2], and Jianping Fu[1,3,4 *]

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

*Correspondence and requests for materials should be addressed to J.F. (email: jpfu@umich.edu).

Keywords: Human pluripotent stem cells; Synthetic embryology; Machine learning; Image processing

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 lumenal 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 lumenal 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.

40 **Insight, innovation, integration**

41 Human pluripotent stem cells (hPSCs) have been successfully used to model and understand
42 cellular events that take place during human embryogenesis. Understanding how cell-cell and
43 cell-environment interactions guide cell actions within a hPSC-based embryo model is a key step
44 in elucidating the mechanisms driving system-level embryonic patterning and growth. In this
45 work, we present a robust video analysis pipeline that incorporates the use of machine learning
46 methods to fully characterize the process of hPSC self-organization into lumenal cysts to mimic
47 the lumenal epiblast cyst formation soon after implantation of the human blastocyst. This
48 pipeline will be a useful tool for understanding cellular mechanisms underlying key embryogenic
49 events in embryo models.

50

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 trophectoderm 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 trophectoderm cells into the uterine wall and their differentiation into cytotrophoblast and syncytiotrophoblasts as well as the development of the epiblast into a lumenal rosette structure enclosing a central cavity. Soon after, the epiblast cells next to invading trophectoderm 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 towards 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

74 the formation of lumenal hPSC cysts and the differentiation of hPSCs into amnion cells. Shao *et*
75 *al.*[13]were the first to show that hPSCs could differentiate into amnion cells. They engineered a
76 3D biomimetic platform with a soft gel bed made with the basement membrane matrix Geltrex™
77 and a 3D matrix overlay made with a low concentration of Geltrex™ diluted in culture medium.
78 In this system, hPSC clusters would undergo lumenogenesis and form lumenal structures
79 containing a central cavity. Over time, three types of cysts resulted from lumenal hPSC clusters:
80 cyst composed of amnion cells, cysts composed of undifferentiated hPSCs, and asymmetric cysts
81 containing amniotic cells at one pole and undifferentiated hPSCs at the opposite pole (Fig. 1).
82 The percentage of each type of cyst was shown to depend heavily on the initial cell plating
83 density. While BMP-SMAD signaling was found to be important for amnion differentiation, the
84 mechanism(s) that led to the initiation of amnion differentiation in the 3D structure has remained
85 elusive.

86 The ability of hPSCs to self-organize and differentiate into *in vivo*-like structures in *in*
87 *vitro* settings posits the existence of endogenous developmental programs. Consequently, a
88 crucial characteristic of *in vivo*-relevant, stem cell-based embryo models is their ability to
89 leverage these programs in order to capture the progressive nature of human development.
90 Triggering these developmental programs, however, is not a trivial endeavor; it requires cell
91 culture environments engineered with correct dimensionality as well as correct mechanical and
92 biochemical properties. Having taken the necessary first step of creating a hPSC model that
93 recapitulates a developmental period of interest, the next step becomes the elucidation of the
94 mechanisms at work in the system. *In vivo*, progressive development entails branching of distinct
95 lineages and progressive differentiation into cell types with increasingly restricted potential[16].
96 Studying these processes in hPSC models in a tractable manner requires the use of computational

97 tools that minimize manual curation and bias. Machine learning tools have come to the forefront
98 and are increasingly used to parse the mechanisms at work in these systems. To date, however,
99 many of these efforts have been directed towards the application of single-cell RNA-sequencing
100 (scRNA-seq) data analysis tools[17–21] or examining global features of the observed structures at
101 discrete time points[22,23]. While these approaches are useful, their discrete nature limits their use
102 for understanding how factors in the local cell microenvironment trigger and guide cell state
103 changes that lead to the emergence of relevant structures. Understanding this requires the ability
104 to continuously monitor individual cells in the system and record division events for later lineage
105 tracing.

106 There have been several efforts towards the creation of classifiers for the identification of
107 dividing cells. While the methods are varied, the models can be divided into two categories: (1)
108 models that use spatial features[24–26] and (2) models that use both spatial and temporal features[27–
109 33]. Many of the models that rely only on spatial features for classification utilize morphological
110 feature extraction that leverages the clear differences in visual characteristics between dividing
111 and non-dividing cells[24–26]. Models that lack temporal information, however, face the additional
112 challenge of having to consider how the timing at which the event is captured will affect the
113 features of interest. This is not an issue for spatiotemporal models in which many stages of the
114 division process can be captured and used for the classification. However, rather than focusing
115 on the nucleus, which shows the most obvious visual changes during division, many of the
116 existing models rely on phase contrast microscopy images[27–30,32,33] that complicate classification
117 because of the confounding factor of varying cell shape. In this work, we present a
118 computational tool for the comprehensive analysis of live cell imaging data of hPSC cyst
119 formation using a unique nuclear GFP H9 hESC reporter line. Using Python, we created a

120 pipeline that is able to process the images and identify all individual cells in a developing hPSC
121 cyst. The pipeline captures information on both the cell properties and cell neighborhood at each
122 time point. Further, we trained a machine learning model for event recognition that is able to
123 identify changes in cell state such as division and death by looking at spatiotemporal properties
124 of the nuclei. With this tool, we hope to parse the relationship between the properties of the local
125 environment and cell-level decisions that lead to emergent behaviors like hPSC cyst formation
126 and growth.
127

128 **Materials and Methods**

129 **Cell culture substrate preparation**

130 An array of 100 μm-diameter circular adhesive islands was created using a two-step
131 micropatterning method as described previously[34]. Briefly, a poly-dimethylsiloxane (PDMS)
132 elastomeric stamp with an array of circular posts was generated using replica molding from a
133 silicon mold fabricated by standard photolithography and deep reactive ion etching (DRIE)[35,36].
134 The center-to-center spacing between adjacent posts on the PDMS stamp was 150 μm, and the
135 post height and diameter were 30 μm and 100 μm, respectively. The PDMS stamp was coated in
136 1% Geltrex (Thermo Fisher Scientific; derived from Engelbreth- HolmSwarm tumors similarly
137 as Matrigel®) solution for 24 h at 4 °C and subsequently rinsed with distilled water and blown
138 dry with nitrogen. Before stamping, the cell culture substrate was prepared by coating a glass
139 coverslip with PDMS and treating it with ultraviolet (UV) ozone (UV-ozone cleaner; Jelight,
140 Irvine, CA) for 7 min to oxidize the PDMS surface. The PDMS stamp was then placed in
141 conformal contact with the PDMS-coated coverslip for 5 s to transfer Geltrex from the stamp to
142 the coverslip. To restrict cell attachment to the circular adhesive islands, the coverslip was

143 treated with Pluronic F127 NF dissolved in PBS (0.2%, *w* / *v*; BASF, Ludwigshafen, Germany)
144 for 1 h at room temperature and rinsed with distilled water. The coverslip was then immersed in
145 mTeSR (STEMCELL Technologies) for a minimum of 2 h to further block the non-
146 functionalized surface of the coverslip. Finally, the coverslip was submerged in mTeSR medium
147 containing 1% Geltrex for 1 h. The coverslip was washed with PBS before cell seeding.
148

149 **Cell culture**

150 H9 hESCs (WA09, WiCell; NIH registration number: 0062) were used in this study. All culture
151 protocols have been pre-approved by the Human Pluripotent Stem Cell Research Oversight
152 Committee at the University of Michigan. The H9 hESC line is authenticated by the original
153 source, and further authenticated in house by immunostaining for pluripotency markers and
154 differentiation into the three germ layers. Karyotype analysis was performed by Cell Line
155 Genetics. The H9 hESC line was tested negative for mycoplasma contamination (LookOut
156 Mycoplasma PCR Detection Kit, Sigma-Aldrich). H9 cells were cultured in a feeder-free culture
157 system using mTeSR medium and lactate dehydrogenase-elevating virus (LDEV)-free, hESC-
158 qualified reduced growth factor basement membrane matrix Geltrex, per manufacturer's
159 instructions. During each passage, cell culture was visually examined to remove spontaneously
160 differentiated, mesenchymal-like cells. All hESCs used in this work had passage numbers less
161 than P70.
162

163 **Generation of mTnG cells**

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

166 (Addgene plasmid #32610). The PCR product was then ligated into the ePiggyBac vector with a
167 constitutively active puromycin selection cassette[37]. membrane-tdTomato was PCR amplified
168 from a gift plasmid pQC membrane TdTomato IX (Addgene plasmid #37351). The PCR product
169 was then ligated into the ePiggyBac vector with a constitutively active neomycin selection
170 cassette[37]. These two plasmids (1.5 μg each) were co-transfected with 1 μg pCAG-PBase
171 (ePiggyBac transposase helper plasmid obtained from Dr Ali H. Brivanlou[37]) using GeneJammer
172 (Agilent Technologies) into H9 hESCs that were plated at 50,000 cells cm^{-2} 24 h prior to
173 transfection. Puromycin selection (2 μg mL^{-1}) and G418 selection (250 μg mL^{-1}) started at 4 days
174 after transfection. The cells were selected for 7 days. After selection, the cells were dissociated to
175 single cells and replated at low density (400 cells cm^{-2}) for clone picking. 12 clones were hand-
176 picked and evaluated for brightness and pluripotency. 3 clones were expanded at the end (mTnG
177 #1, 2, 3). mTnG #1 hESC line has the brightest fluorescent signal and is used in the current
178 study.
179
180 **Cyst formation assay**
181 Cultured hESC colonies were dissociated into single cells with Accutase (Sigma-Aldrich) at 37
182 °C for 10 min before the cells were centrifuged and re-suspended in mTeSR1 medium containing
183 10 μM ROCK inhibitor, Y27632 (Tocris), to avoid dissociation-induced apoptosis[38]. Cells were
184 then plated onto coverslips pre-coated with circular adhesive islands at a density of 300,000 cells
185 cm^{-2}. To establish 3D ECM overlay, culture medium was changed to fresh mTeSR1 medium
186 containing 10 μM Y27632 and 4% (*v* / *v*) Geltrex 2 h after initial cell seeding. Y27632 was
187 removed 24 h after initial cell seeding, at which time the coverslip was transferred to
188 fluorescence microcopy for live cell imaging.

189

190 **Live cell video acquisition**

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

197

198 **Image pre-processing**

199 A customized Python program was used to process raw images collected from live cell imaging
200 using fluorescence microscopy. First, contrast was enhanced using adaptive image enhancement
201 developed by Peng *et al*[39]. Specifically, each pixel in the image is normalized using the mean and
202 variance of a local region surrounding the pixel. This local region is determined adaptively. For a
203 given pixel, the program starts from a given initial size and expands until the standard deviation
204 of the region is equal to or more than a given threshold. For computational tractability, maximum
205 radius was set at 5 pixels. The threshold is in the range of 0.2 to 0.8 and is meant to ensure that
206 the local region has enough relevant structures to classify a pixel as being part of the background
207 or part of an object. After obtaining the region size, the pixel is normalized by subtracting the
208 local mean and dividing by the local standard deviation. This will account for varying
209 background intensity and varying contrast, respectively. A background mask is then obtained by
210 binarizing the resulting image with a binary threshold (cv2.THRESH_BINARY). This
211 background mask is further refined with a dilation (cv2.dilate, kernel size = (3,3), iterations = 2)

212 followed by an erosion (cv2.erode, kernel size = (5,5), iterations = 2). Multiplying this resulting
213 background mask by the original image eliminates background noise. Second, contours of cells
214 were identified with the use of adaptive gaussian threshold (cv2.adaptiveThreshold with
215 blockSize = 23, $C = 1$) (Fig. 3b). For each pre-processed image, contours were extracted with the
216 use of cv2.findContours with cv2.RETR_TREE and cv2.CHAIN_APPROX_NONE. The third
217 step is to carry out segmentation to find individual cell contours (Fig. 3d). The pipeline measures
218 the area and circularity of each contour. Contours identified as individual cells are stored. The
219 contours identified as cell clusters undergo concavity point-pair segmentation, a method
220 developed by Farhan *et al.* based on finding concavity point-pairs using a variable-size
221 rectangular window[40]. In brief, using an established interval, a list of contour coordinates is first
222 extracted from the binary image of the cell cluster. For each coordinate in the list, lines are
223 drawn to the next two points in the list. Once a line passes through the image background (*i.e.*, a
224 pixel with value 0), the algorithm finds the contour coordinate at which the line no longer passes
225 through the background and establishes this coordinate as a concavity point. After filtering the
226 resulting point list to account for contour irregularities, the program finds the directionality
227 vector of each concave area. Using this vector, each concavity point establishes a rectangular
228 window in which to search for other concavity points. Once all concavity points have paired up,
229 a line is drawn between them and the cluster is segmented. Farah *et al.* validated the method with
230 the use of three data sets, two of which contained bright field microscopy images of yeast cells,
231 and one which contained fluorescent microscopy images of yeast cells[40]. They showed that the
232 concavity point-pairing segmentation method was highly effective, with precision averaging at
233 0.98[40].
234

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 (Figure 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

258 randomly shuffled and images were transformed using rotation (range of 5), shifts in height
259 (range of 0.1) and shifts in width (range of 0.1). Transformations in this data generator class
260 were carried out using the Keras ImageDataGenerator class. One-hot encoding was applied to the
261 labels before training.
262
263 **Video analysis pipeline**
264 Live cell videos were analyzed with a Python pipeline. Images were pre-processed and all
265 individual nuclei were identified. A cell tracker python class was used to give each cell a unique
266 identification (ID) number and track cells from one time point to another using Euclidean
267 distance. The nuclei were cropped from the image and stored in a Python dictionary. For event
268 recognition, the cropped nuclei of the current time point and the cropped nuclei from the
269 previous two time points were passed as input to the CNN-LSTM classifier. Whenever a new
270 cell would appear in the environment, the parent cell would be identified using a parent score
271 (ps) parameter. Newly divided cells tend to be small, bright, and similar in size. For this reason,
272 the ps takes into account both the classification of the cells in the previous time point and the
273 similarity in area and brightness of the nuclei between the new cell and the possible sister cell.
274 The cell with the highest ps in the local neighborhood of the newly appeared cell would be
275 assigned as the parent. At this point, the daughter cell with the parent ID receives a new ID, and
276 the parent ID of the daughter cells is stored. The number of neighboring cells and the average
277 distance to neighbors are stored for each cell at every time point. The output of the video cell
278 analysis is a Python DataFrame with the cell IDs, cell positions, number of neighbors, average
279 distance to neighbors and parent IDs for each time point.
280

281 **Results**

282 **Cyst formation analysis pipeline**

283 Analysis of the morphogenesis of a multicellular structure at discrete time points can give
284 insight into the system-level dynamics governing its growth and development. However,
285 important cell-level dynamics and the degree of stochasticity and heterogeneity in a multicellular
286 system remain difficult to elucidate. To better study the dynamic process of hPSC cyst
287 formation, we developed an experimental platform to provide a biomimetic niche for the
288 formation of hESC cysts in a controllable and robust manner. Specifically, an array of 100 μm-
289 diameter circular adhesive islands was created on a coverslip, before hESCs expressing
290 membrane tdTomato and nucleus-EGFP (mTnG) were seeded onto the coverslip. Two hours
291 after cell seeding, culture medium was changed to fresh mTeSR1 medium containing 10 μM
292 Y27632 and 4% (*v* / *v*) Geltrex, to establish a 3D ECM overlay. The coverslip was transferred to
293 a Zeiss Axio Observer Z1 inverted epifluorescence microscope 24 h after cell seeding. To track
294 the dynamics of hESC cyst formation, live cell imaging was conducted for 24 h with a depth of
295 focus that captured all the cells in the system. Given their self-organizing property, hESCs
296 confined on adhesive islands on the coverslip formed small clusters and underwent epithelization
297 and lumenogenesis. Throughout cyst formation, cells showed limited movement ability in the z-
298 direction. Live cell imaging data were then processed with the use of a Python pipeline capable
299 of image processing, cell tracking, and event recognition. With these extracted data, a
300 comprehensive characterization of cell states and actions during hESC cyst formation could be
301 conducted, using a workflow that includes image processing with machine learning, and the
302 characterization of cell state and cyst growth (Fig. 2).
303

304 **Event recognition and image processing**

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

320 Having successfully trained a machine learning model for event recognition, we next
321 sought to create an image processing pipeline capable of identifying individual cell nuclei. While
322 there are a number of computational tools available for identification and tracking of cells in a
323 multicellular system[41–43], identifying and tracking cells within a forming cyst presents unique
324 challenges that require the use of more catered approaches. For example, in our test of the
325 commonly used watershed method for segmentation, it is difficult to carry out correct
326 segmentation of cell clusters, likely because of the compact arrangement of cells in hESC cysts.

327 Additionally, global thresholding methods were unsatisfactory for finding hESC clusters, likely
328 because of variations in nuclear GFP intensity. To address these challenges, we devised an
329 imaging processing pipeline uniquely suited to carry out thresholding and segmentation in tightly
330 packed hESC cysts. The image processing steps consist of thresholding, denoising,
331 segmentation, and identification of individual cells in hESC cysts (Fig. 4). For the segmentation
332 step, we utilized adaptive local enhancement[39] to enhance the contrast between nuclei and
333 background. For binarization, we utilized adaptive gaussian thresholding to ensure that cells that
334 are slightly out of focus can still be identified. After binarization and denoising, hESC clusters
335 were segmented with the use of concavity point pairing analysis[40]. Cells are given a unique
336 identification (ID) number and tracked from one time point to another with the use of Euclidean
337 distance.
338
339 **Live cell data processing and system characterization**
340 As mentioned earlier, the experimental platform for the formation of hESC cysts consisted of an
341 array of micropatterned Geltrex islands with a diameter of 100 μm generated with a two-step
342 micropatterning process[34] (Fig. 5a). After a period of 24 h in which mTnG hESCs were allowed
343 to attach and cluster onto the adhesive islands, the experimental platform was transferred to a
344 fluorescent microscope for live cell imaging for a period of 24 h (Fig. 5b&c). To avoid cytotoxic
345 effects, images were recorded at intervals of 10 min.
346 Having obtained live cell data from various hESC cysts, we processed the images using
347 the Python pipeline. Figure 6 shows the growth profiles of four different hESC cysts. The growth
348 profiles vary greatly between the cysts. While there are periods of a sustained increase in cell
349 number like the one seen in Figure 6a between 500 min and 750 min, we can also find periods of

350 a sustained decrease in the number of cells as seen in Figure 6c between 0 min and 250 min.
351 Regardless of the growth profile, however, the number of cells seem to plateau for all the cysts.
352 While the final cell numbers might be similar among the four hESC cysts, there are a number of
353 different growth trajectories that could not have been inferred from looking at the final cyst
354 configuration. In the context of modeling human development, this information facilitates the
355 study of how these changes in growth dynamics correspond to relevant cell specification events.
356 Further, as cells progressively differentiate and more populations appear in a system, we can start
357 to study the growth dynamics of specific populations and how they relate to correct form and
358 function in the structure.

359 Lineage tracing is a powerful tool for parsing the mechanisms guiding morphogenetic
360 events in a multicellular system. It has many uses including providing insight into the timing of
361 differentiation of cell types of interest and helping identify lineage-specific precursor cells.
362 Combined with the ability to record properties of the local cell microenvironment, it can help
363 parse when and why different cell types arise. Figure 7a shows a network representation of the
364 cyst shown in Fig. 6a at different time points. Edges between cells of a given cyst are connected
365 to each other based on an established threshold distance. The solid line going from one time
366 point to another indicates a chosen cell lineage, with an additional dotted line indicating cell
367 divisions. Having tracked a cell and established its lineage, we can characterize the local cell
368 density experienced by the cells in the lineage throughout time. As can be seen in Figure 7b, the
369 number of cells in the neighborhood and the average distance from neighbors of the selected cell
370 continue to vary even after cell number in the cyst has plateaued (Figure 6a). We can also see
371 from Figure 7d that the cyst radius continues to increase after cell number has plateaued,
372 suggesting that structure growth does not necessarily correlate with increased cell number. From

373 the MSD plot in Figure 7d we can see that the average MSD in the system remains low
374 throughout the 24 hours. This is likely a result of the confinement provided by the adhesive
375 islands in which the cells exist. As different cell types start to arise, this pipeline output can be
376 used to assess differences in movement dynamics between different cell populations and their
377 spatial segregations. For example, mesoderm cells, which are studied in our post-implantation
378 amniotic sac embryoid (PASE) model[13], are known to be more migratory as compared to other
379 populations like epithelial ectoderm and endoderm cells. While the model presented here is
380 limited to the first 24 h of cyst formation, future efforts can be devoted to extending this
381 timeframe to include important events such as symmetry breaking caused by the appearance of
382 amnion-like cells (AMLCs) in the PASE. In our PASE model, we found that the initial cell
383 seeding density has a significant effect on morphogenesis and differentiation of hESCs[13,34]. With
384 the information that can be obtained from this pipeline, we can begin to understand the role of
385 initial conditions like cell seeding density, and we can begin the work of relating changes in the
386 local environment with cell-level decisions that lead to the cyst-level growth and patterning.
387

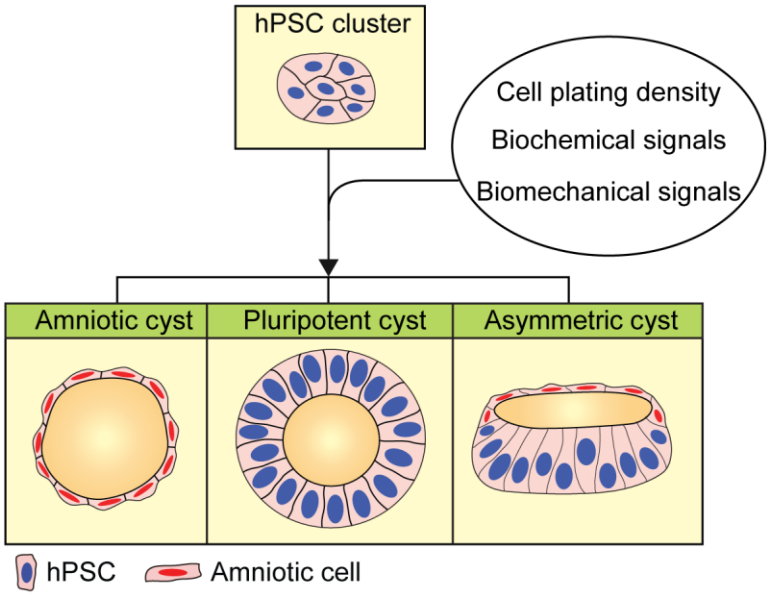
388 **Conclusions**

389 The successful generation of human embryo-like structures is a crucial step in advancing
390 fundamental understanding of human development, without using intact, natural human
391 embryos. However, limitations on the insights gained through analysis of human embryo-like
392 structures at discrete time points drove us to create a live cell video processing pipeline catered
393 for the unique challenges of our system. With the use of both spatial and temporal information,
394 we were able to create a machine learning model for event recognition. Furthermore, this model
395 was integrated into an image processing pipeline that leveraged specialized image processing

396 tools for the identification and tracking of individual cells in our system. With this integrative
397 pipeline we were able to characterize the cell states and actions during the dynamic growth and
398 morphogenesis of lumenal hESC cysts. Combining this tool with reporter lines for cell types of
399 interest, we hope to advance in our goal to elucidate the mechanisms driving lumenogenesis, cyst
400 growth, and cell fate specification in our *in vitro* hESC models of human development.
401
402 **Acknowledgements**
403 A.M.R.I. is partially supported by the National Science Foundation Graduate Research
404 Fellowship under grant no. DGE 1256260. This research is supported by the Michigan-
405 Cambridge Collaboration Initiative, the University of Michigan Mcubed Fund, the 21st Century
406 Jobs Trust Fund received through the Michigan Strategic Fund from the State of Michigan
407 (Grant CASE-315037), the National Institutes of Health (R21 NS113518 and R21 HD100931),
408 and the National Science Foundation (CMMI 1917304 and CBET 1901718).
409
410
411
412
413
414
415

416 **FIGURES AND FIGURE LEGENDS**

417 **Figure 1**



418

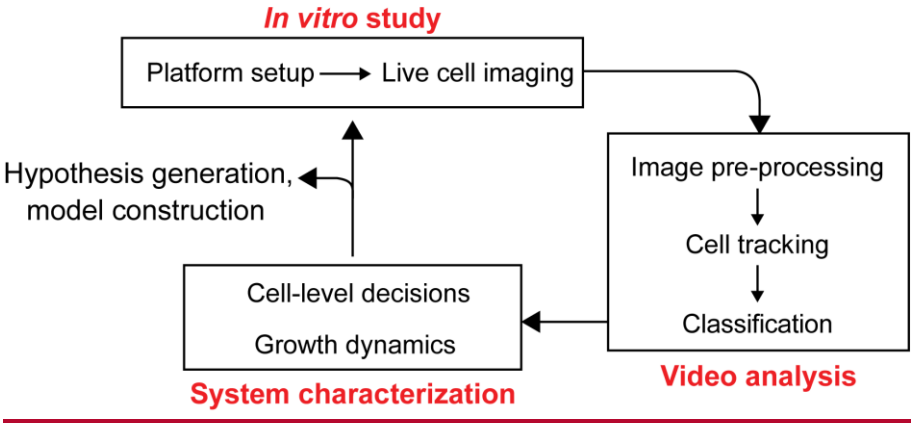
419 **Figure 1.** Development of hPSC clusters into three distinct types of lumenal structures: amniotic

420 cyst, pluripotent cyst, and asymmetric cyst. Amniotic cysts are composed of amniotic cells,

421 whereas pluripotent cysts are composed of undifferentiated hPSCs. Asymmetric cysts contain

422 amniotic cells at one pole and undifferentiated hPSCs at the opposite pole.

423 **Figure 2**



424

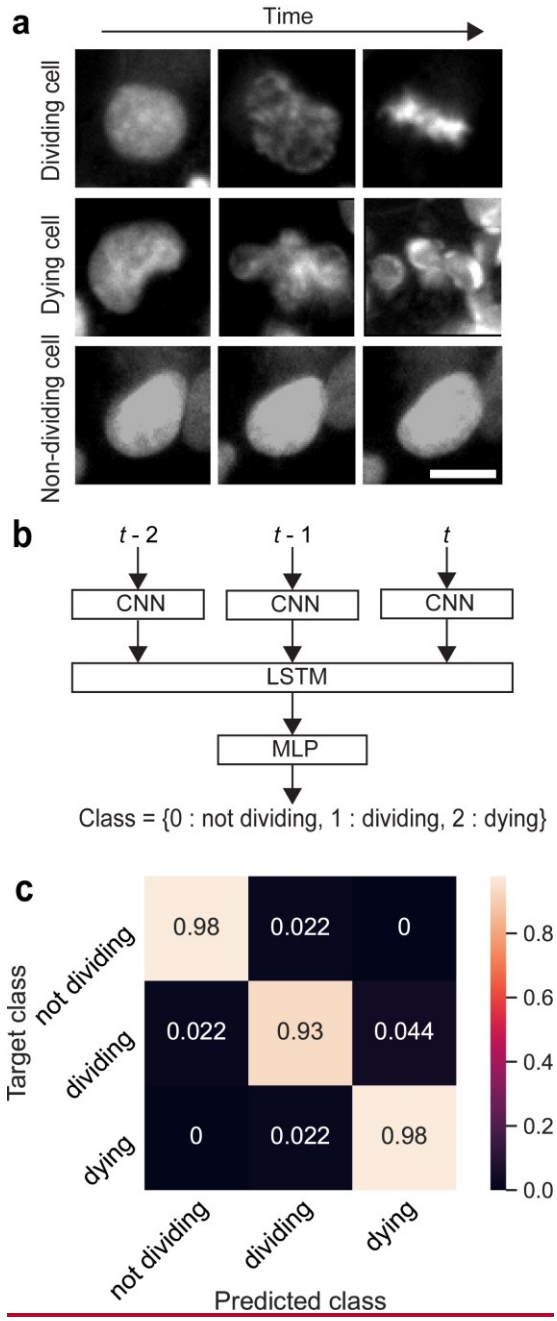
425 **Figure 2.** Workflow for live cell imaging data analysis of cyst formation, including *in vitro*

426 experimentation, video analysis, and system characterization.

427

428

429 **Figure 3**



430

431 **Figure 3.** (**a**) Input image sequences to CNN-LSTM classifier. Images show the GFP channel of

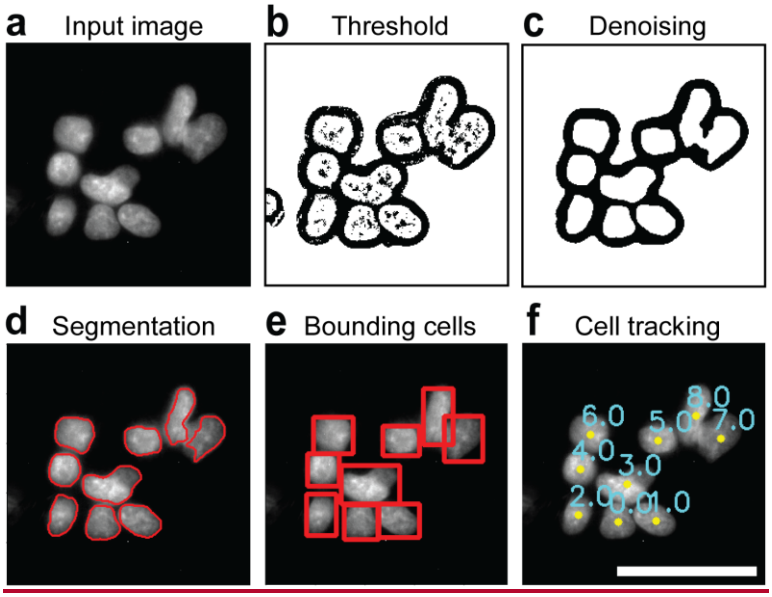
432 imaged mTnG H9 hESCs. The top, middle and bottom rows correspond to a dividing, dying, and

433 non-dividing cell, respectively. Scale bar, 10μm. (**b**) CNN-LSTM framework followed by a

434 multilayer perceptron (MLP) for multiclass classification. (**c**) Confusion matrix for CNN-LSTM

435 classifier.

436 **Figure 4**



437

438 **Figure 4.** Image processing pipeline. Cell cluster is first isolated from original images obtained

439 from live cell imaging of mTnG H9 hESCs (GFP channel is shown here) (**a**), before going

440 through adaptive local enhancement and adaptive Gaussian thresholding (**b**). The image is then

441 denoised with the use of open, erode, and filtering by connectivity (**c**). Cell clusters are

442 segmented with concavity point analysis and individual cell contours are established (**d**).

443 Bounding rectangle is then inputted into cell tracker (**e**), and cells are given a unique ID (**f**).

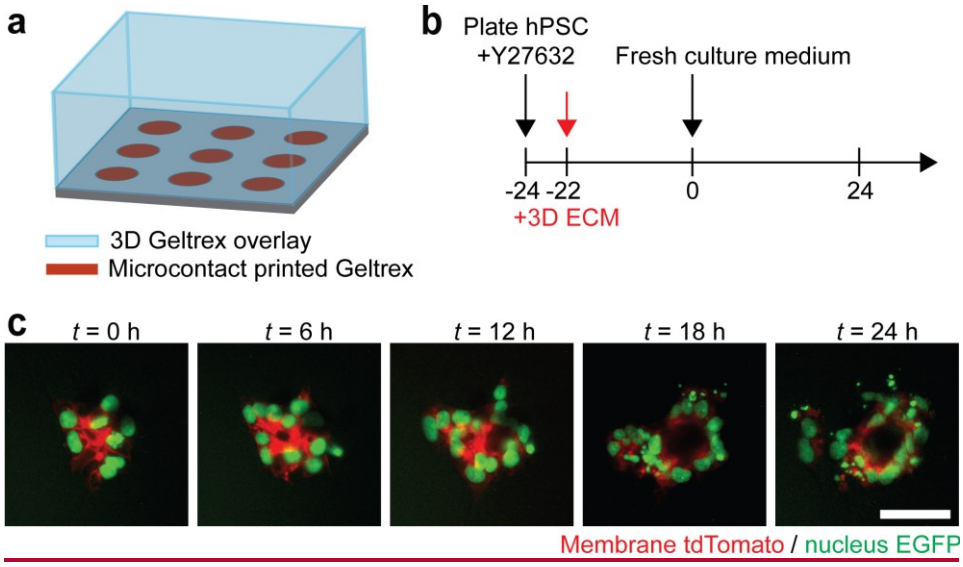
444 Scale bar, 50 µm.

445

446

447

448 **Figure 5**



449

450 **Figure 5.** hPSC cyst formation experiment. (**a**) Cartoon of experimental platform consisting of

451 micropatterned adhesive islands and a 3D Geltrex overlay. (**b**) Experimental protocol timeline.

452 (**c**) Live imaging of a hPSC cluster at different time points. In this assay, the mTnG H9 hESC

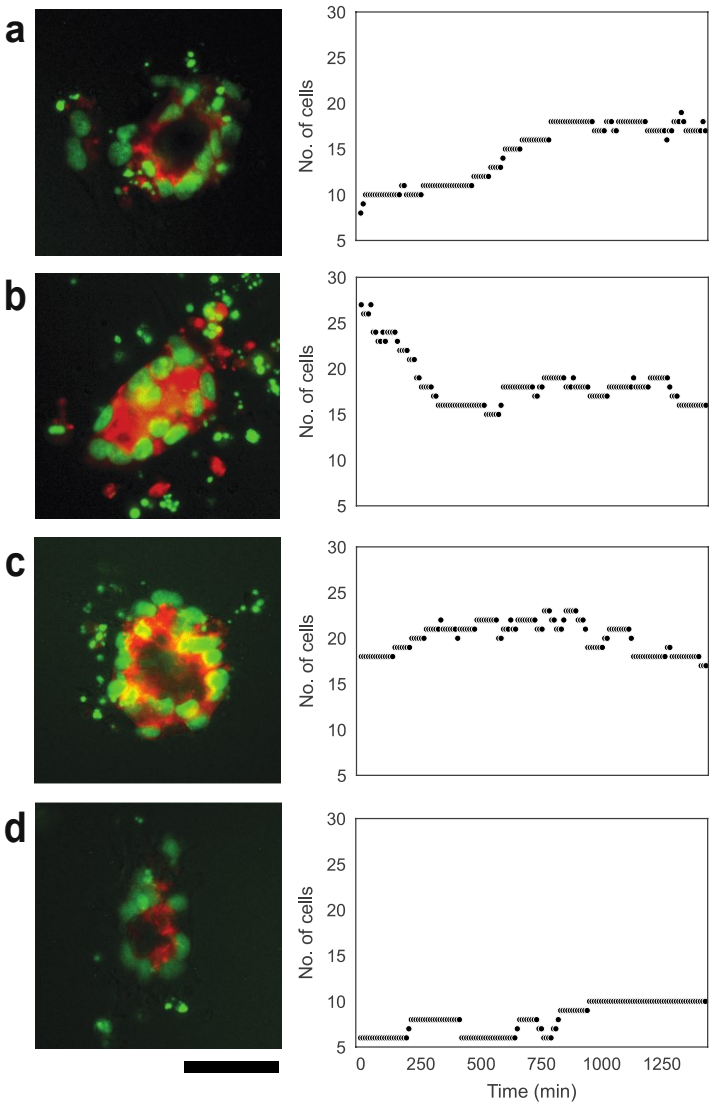
453 line was used. Scale bar, 50 μm.

454

455

456

457 **Figure 6**



458

459 **Figure 6.** Growth dynamics of four different hPSC cysts. Left and right columns show the cysts

460 after 24 hours of growth and the number of cells in the cyst through the period of 24 hours,

461 respectively. In this assay, the mTnG H9 hESC line was used. Images show merged nuclear

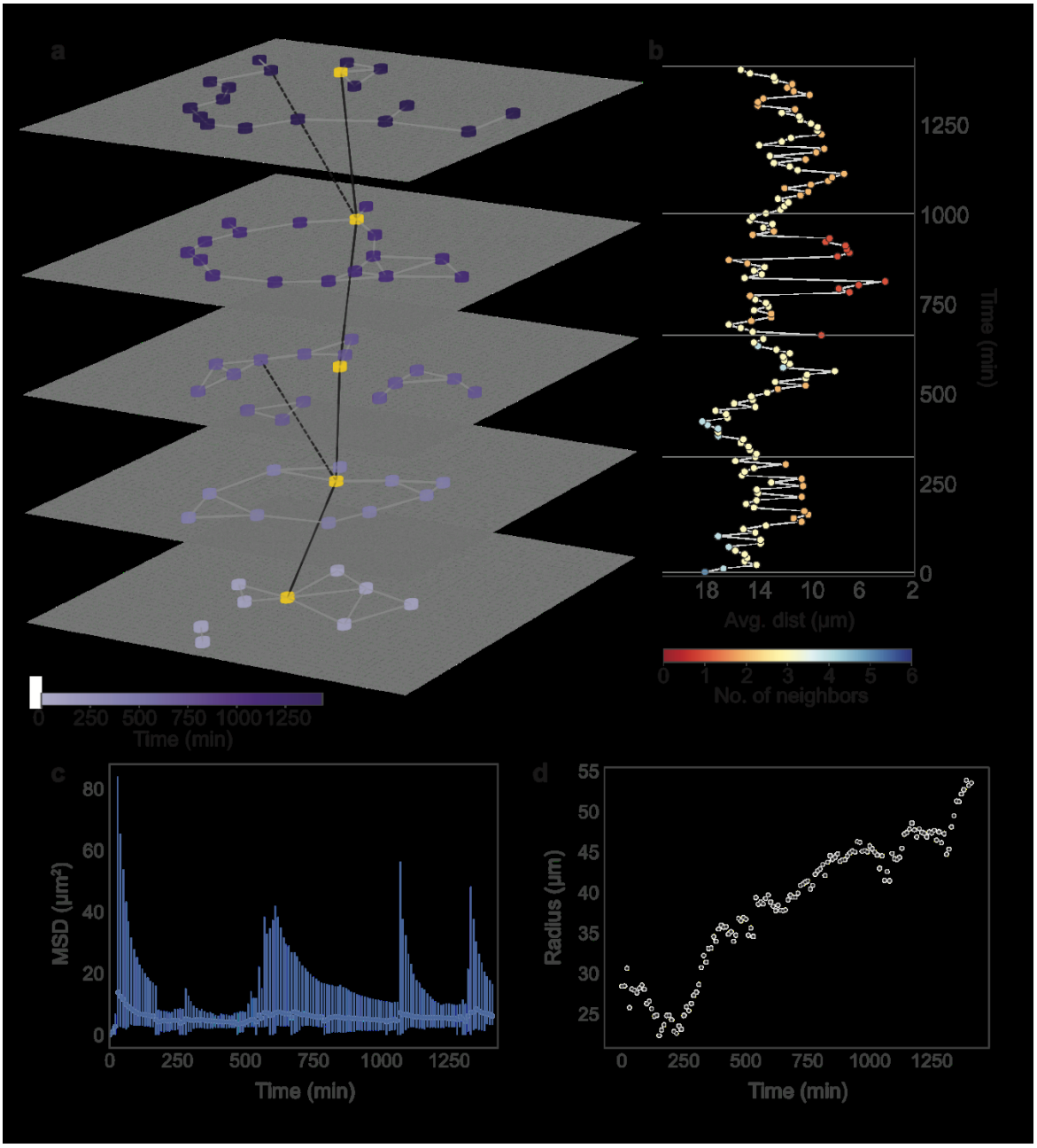
462 EGFP (green) and membrane tdTomato (red). Scale bar, 50 μm.

463

464

465

466 **Figure 7**



467

468 **Figure 7.** Characterization of hPSC cyst formation dynamics. (**a**) Network representation of a

469 cyst at different time points. The color of the nodes in each plane indicates the time point. Each

470 node represents a single cell in the cell cluster. Connections between cells are established based

471 on a threshold distance. Cell lineage of the gold-colored cell is shown with the black line

472 connecting the cell at different time points. Dotted lines indicate instances of cell division. (**b**)

Average distance to neighbors and number of neighbors corresponding to the gold-colored cell in **a** as a function of time. (**c**) Average mean squared displacement (MSD) of all the cells in the cyst as a function of time. 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.

480 **References**

481 1. Pera, M. F. Human embryo research and the 14-day rule. *Dev.* **144**, 1923–1925 (2017).

482 2. Rivron, N. & Pera, M. Debate ethics of embryo models from stem cells. *Nature* **564**, 183–
483 185 (2018).

484 3. Ma, H. *et al.* In vitro culture of cynomolgus monkey embryos beyond early gastrulation.
485 *Science (80-.).* **366**, (2019).

486 4. Niu, Y. *et al.* Dissecting primate early post-implantation development using long-term in
487 vitro embryo culture. *Science (80-.).* **366**, eaaw5754 (2019).

488 5. Sasaki, K. *et al.* The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent
489 Amnion. *Dev. Cell* **39**, 169–185 (2016).

490 6. Sozen, B. *et al.* Self-assembly of embryonic and two extra-embryonic stem cell types into
491 gastrulating embryo-like structures. *Nat. Cell Biol.* **20**, 979–989 (2018).

492 7. Sozen, B. *et al.* Self-Organization of Mouse Stem Cells into an Extended Potential
493 Blastoid. *Dev. Cell* **51**, 698-712.e8 (2019).

494 8. Resto Irizarry, A. M., Nasr Esfahani, S. & Fu, J. Bioengineered pluripotent stem cell
495 models: new approaches to explore early human embryo development. *Curr. Opin.*
496 *Biotechnol.* **66**, 52–58 (2020).

497 9. Taniguchi, K., Heemskerk, I. & Gumucio, D. L. Opening the black box: Stem cell-based
498 modeling of human post-implantation development. *J. Cell Biol.* **218**, 410–421 (2019).

499 10. Shahbazi, M. N., Siggia, E. D. & Zernicka-Goetz, M. Self-organization of stem cells into
500 embryos: A window on early mammalian development. *Science (80-.).* **364**, 948–951
501 (2019).

502 11. Fu, J., Warmflash, A. & Lutolf, M. P. Stem-cell-based embryo models for fundamental

503 research and translation. *Nat. Mater.* **20**, 132–144 (2021).

504 12. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to
505 recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat.*
506 *Methods* **11**, 847–854 (2014).

507 13. Shao, Y. *et al.* A pluripotent stem cell-based model for post-implantation human amniotic
508 sac development. *Nat. Commun.* **8**, 1–15 (2017).

509 14. Simunovic, M. *et al.* A 3D model of a human epiblast reveals BMP4-driven symmetry
510 breaking. *Nat. Cell Biol.* **21**, 900–910 (2019).

511 15. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using
512 stem cells. *Nature* **573**, 421–425 (2019).

513 16. Blanpain, C. & Simons, B. D. Unravelling stem cell dynamics by lineage tracing. *Nat.*
514 *Rev. Mol. Cell Biol.* **14**, 489–502 (2013).

515 17. Chen, L. *et al.* Investigating the gene expression profiles of cells in seven embryonic
516 stages with machine learning algorithms. *Genomics* **112**, 2524–2534 (2020).

517 18. Shah, N. *et al.* An Experiment on Ab Initio Discovery of Biological Knowledge from
518 scRNA-Seq Data Using Machine Learning. *Patterns* **1**, 100071 (2020).

519 19. Peng, L. *et al.* Single-cell RNA-seq clustering: datasets, models, and algorithms. *RNA*
520 *Biol.* **17**, 765–783 (2020).

521 20. Andrews, T. S. & Hemberg, M. Identifying cell populations with scRNASeq. *Mol.*
522 *Aspects Med.* **59**, 114–122 (2018).

523 21. Liu, J., Fan, Z., Zhao, W. & Zhou, X. Machine Intelligence in Single-Cell Data Analysis:
524 Advances and New Challenges. *Front. Genet.* **12**, (2021).

525 22. Guo, J. *et al.* Machine learning-assisted high-content analysis of pluripotent stem cell-

526 derived embryos in vitro. *Stem Cell Reports* **16**, 1331–1346 (2021).

527 23. Chen, K. *et al.* Branching development of early post-implantation human embryonic-like
528 tissues in 3D stem cell culture. *Biomaterials* **275**, 120898 (2021).

529 24. Irshad, H. *et al.* Automated mitosis detection using texture, SIFT features and HMAX
530 biologically inspired approach. *J. Pathol. Inform.* **4**, 12 (2013).

531 25. Liu, A., Hao, T., Gao, Z., Su, Y. & Yang, Z. Nonnegative mixed-norm convex
532 optimization for mitotic cell detection in phase contrast microscopy. *Comput. Math.*
533 *Methods Med.* **2013**, (2013).

534 26. Huang, C. H. & Lee, H. K. Automated mitosis detection based on eXclusive Independent
535 Component Analysis. *Proc. - Int. Conf. Pattern Recognit.* 1856–1859 (2012).

536 27. Yang, F., Mackey, M. A., Ianzini, F., Gallardo, G. & Sonka, M. Cell segmentation,
537 tracking, and mitosis detection using temporal context. *Lect. Notes Comput. Sci.*
538 *(including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **3749 LNCS**, 302–
539 309 (2005).

540 28. Gallardo, G. M., Yang, F., Ianzini, F., Mackey, M. & Sonka, M. Mitotic cell recognition
541 with hidden Markov models. *Med. Imaging 2004 Vis. Image-Guided Proced. Disp.* **5367**,
542 661 (2004).

543 29. Liu, A. A., Li, K. & Kanade, T. Mitosis sequence detection using hidden conditional
544 random fields. *2010 7th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI*
545 *2010 - Proc.* 580–583 (2010) doi:10.1109/ISBI.2010.5490279.

546 30. Su, Y. T., Lu, Y., Chen, M. & Liu, A. A. Spatiotemporal Joint Mitosis Detection Using
547 CNN-LSTM Network in Time-Lapse Phase Contrast Microscopy Images. *IEEE Access* **5**,
548 18033–18041 (2017).

549 31. Harder, N. *et al.* Automatic analysis of dividing cells in live cell movies to detect mitotic
550 delays and correlate phenotypes in time. *Genome Res.* **19**, 2113–2124 (2009).
551 32. Nie, W. Z., Li, W. H., Liu, A. A., Hao, T. & Su, Y. T. 3D Convolutional Networks-Based
552 Mitotic Event Detection in Time-Lapse Phase Contrast Microscopy Image Sequences of
553 Stem Cell Populations. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*
554 1359–1366 (2016) doi:10.1109/CVPRW.2016.171.
555 33. Li, K. *et al.* Computer vision tracking of stemness. *2008 5th IEEE Int. Symp. Biomed.*
556 *Imaging From Nano to Macro, Proceedings, ISBI* 847–850 (2008)
557 doi:10.1109/ISBI.2008.4541129.
558 34. Nasr Esfahani, S. *et al.* Microengineered human amniotic ectoderm tissue array for high-
559 content developmental phenotyping. *Biomaterials* **216**, 119244 (2019).
560 35. Weng, S., Shao, Y., Chen, W. & Fu, J. Mechanosensitive subcellular rheostasis drives
561 emergent single-cell mechanical homeostasis. *Nat. Mater.* **15**, 961–967 (2016).
562 36. Fu, J. *et al.* Mechanical regulation of cell function with geometrically modulated
563 elastomeric substrates. *Nat. Methods* **7**, 733–736 (2010).
564 37. Lacoste, A., Berenshteyn, F. & Brivanlou, A. H. An Efficient and Reversible
565 Transposable System for Gene Delivery and Lineage-Specific Differentiation in Human
566 Embryonic Stem Cells. *Cell Stem Cell* **5**, 332–342 (2009).
567 38. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic
568 stem cells. *Nat. Biotechnol.* **25**, 681–686 (2007).
569 39. Peng, J. Y., Hsu, C. N. & Lin, C. C. Adaptive image enhancement for fluorescence
570 microscopy. *Proc. - Int. Conf. Technol. Appl. Artif. Intell. TAAI 2010* 9–16 (2010)
571 doi:10.1109/TAAI.2010.13.

572 40. Farhan, M., Yli-Harja, O. & Niemistö, A. A novel method for splitting clumps of convex
573 objects incorporating image intensity and using rectangular window-based concavity
574 point-pair search. *Pattern Recognit.* **46**, 741–751 (2013).
575 41. Schindelin, J. *et al.* Fiji - an Open platform for biological image analysis. *Nat. Methods* **9**,
576 (2009).
577 42. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS*
578 *Biol.* **16**, 1–17 (2018).
579 43. ICY : A NEW OPEN-SOURCE COMMUNITY IMAGE PROCESSING SOFTWARE
580 Unité d ' analyse d ' images quantitative – CNRS URA 2582 – Institut Pasteur - France.
581 *Development* 234–237 (2011).
582
583