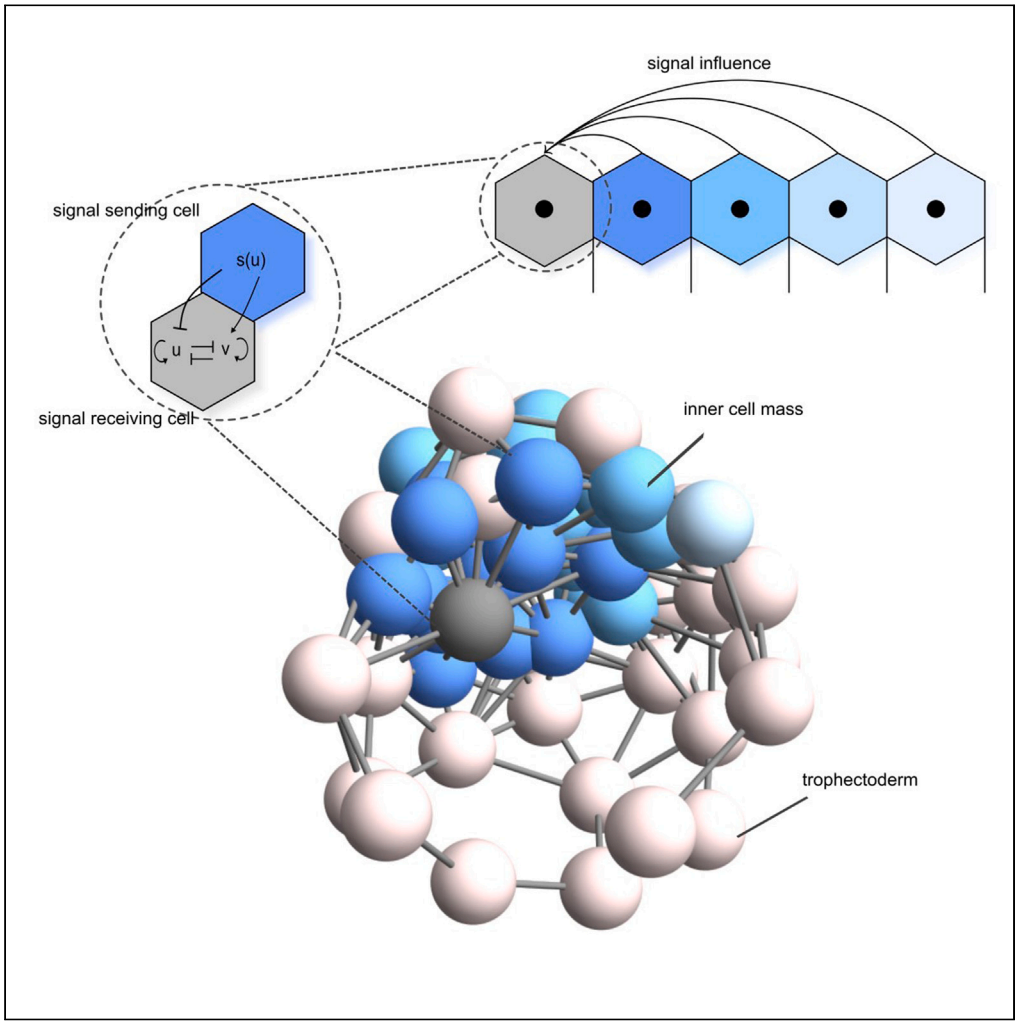
iScience

CellPress
OPEN ACCESS

Article

The salt-and-pepper pattern in mouse blastocysts is compatible with signaling beyond the nearest neighbors



Sabine C. Fischer, Simon Schardt, Joaquín Lilao-Garzón, Silvia Muñoz-Descalzo

sabine.fischer@uni-wuerzburg.de (S.C.F.)
silvia.munoz@ulpgc.es (S.M.-D.)

Highlights

The local cell neighborhood and global ICM population composition correlate

ICM cells show characteristics of local clustering in early and mid mouse blastocysts

ICM patterning requires integration of signals from cells beyond the first neighbors

Fischer et al., iScience *26*, 108106
November 17, 2023
https://doi.org/10.1016/j.isci.2023.108106

iScience

CellPress
OPEN ACCESS

Article

The salt-and-pepper pattern in mouse blastocysts is compatible with signaling beyond the nearest neighbors

Sabine C. Fischer,[1,*] Simon Schardt,[1] Joaquín Lilao-Garzón,[2,3] and Silvia Muñoz-Descalzo[2,4,*]

SUMMARY

Embryos develop in a concerted sequence of spatiotemporal arrangements of cells. In the preimplantation mouse embryo, the distribution of the cells in the inner cell mass evolves from a salt-and-pepper pattern to spatial segregation of two distinct cell types. The exact properties of the salt-and-pepper pattern have not been analyzed so far. We investigate the spatiotemporal distribution of NANOG- and GATA6-expressing cells in the ICM of the mouse blastocysts with quantitative three-dimensional single-cell-based neighborhood analyses. A combination of spatial statistics and agent-based modeling reveals that the cell fate distribution follows a local clustering pattern. Using ordinary differential equations modeling, we show that this pattern can be established by a distance-based signaling mechanism enabling cells to integrate information from the whole inner cell mass into their cell fate decision. Our work highlights the importance of longer-range signaling to ensure coordinated decisions in groups of cells to successfully build embryos.

INTRODUCTION

During the development of multicellular organisms, coordinated cell behavior is essential to shape the embryo. Cellular differentiation, proliferation, and rearrangement is precisely regulated to obtain the correct positioning of the cells. This results in a variety of stable and reproducible patterns across the animal kingdom.[1–3]

The mammalian preimplantation embryo exhibits distinct shape changes and patterns, some of which are conserved among different groups of Eutheria.[4] After fertilization, several cell divisions without cell growth yield a collection of smaller and smaller cells.[5,6] Upon compaction, this loose cellular aggregate transforms into a compact, ball-like structure, and the embryo is referred to as a morula.[4] At the same time, the outside cells of the morula develop an apical-basal polarity. The outside cells gradually form a fluid-tight seal and develop into an epithelium. This marks the first cell fate decision into the inner cell mass (ICM) that is enveloped by the outer trophectoderm (TE) that after implantation gives rise to the embryonic part of the placenta. Subsequently, inside the TE, next to the ICM, a fluid-filled cavity arises and the whole structure is known as blastocyst.

The cells in the ICM differentiate into epiblast (Epi) and primitive endoderm (PrE) precursors that after implantation give rise to the embryo and the yolk sac, respectively.[4] The main markers used for the two lineages are the transcription factors NANOG for the Epi and GATA6 for the PrE.[7] In early blastocysts, ICM cells co-express these markers. During differentiation in mid blastocysts, cells asynchronously downregulate NANOG or GATA6 to form PrE or Epi, respectively.[7]

Besides marker expression, there is also cell positional changes.[8–10] At the late blastocyst stage, the two lineages segregate such that the PrE forms a second epithelium at the surface of the ICM, separating the Epi cells from the cavity. The main pathway involved in the process is FGF/MAPK signaling, which reinforces PrE commitment: epiblast progenitors secrete FGF4, which binds to FGFR1 on Epi and FGFR1 and FGFR2 on PrE-biased cells.[11–17] New analysis tools have been developed to study the neighborhood structure in blastocysts such as IVEN.[6] They reveal that an ICM cell has approximately 30% more neighbors than cells from the surrounding TE. Furthermore, the two cell types are clearly separated such that ICM cells have a majority of ICM neighbors and TE cells have mainly TE neighbors.

To investigate the Epi versus PrE differentiation process, we have previously used three-dimensional imaging of the mouse embryo with single-cell resolution based on immunofluorescence staining, together with three-dimensional neighborhood analyses.[18] This approach enabled studying cell fate acquisition considering single-cell protein expression levels together with cell positional information. Our work

[1]Julius-Maximilians-Universität Würzburg, Faculty of Biology, Center for Computational and Theoretical Biology, Klara-Oppenheimer-Weg 32, Campus Hubland Nord, 97074 Würzburg, Germany
[2]Instituto Universitario de Investigaciones Biomédicas y Sanitarias (IUIBS), Universidad Las Palmas de Gran Canaria (ULPGC), Paseo Blas Cabrera Felipe "Físico" 17, Las Palmas de Gran Canaria 35016, Spain
[3]Present address: Janelia Research Campus, Howard Hughes Medical Institute, Ashburn VA 20147, United States
[4]Lead contact
*Correspondence: sabine.fischer@uni-wuerzburg.de (S.C.F.), silvia.munoz@ulpgc.es (S.M.-D.)
https://doi.org/10.1016/j.isci.2023.108106

Table 1. Overview of datasets and processing steps

Dataset	Reference	Imaging	Image analysis	Population assignment	Neighborhood extraction	No of embryos
I	New embryo data	This paper	See Fischer et al.[18]	This paper	This paper	248
II	Fischer et al.[18]	See Fischer et al.[18]	See Fischer et al.[18]	This paper	This paper	44
III	Saiz et al.[7]	See Saiz et al.[7]	See Saiz et al.[7]	See Saiz et al.[7]	This paper	137
IV	Saiz et al.,[22] stained with anti-NANOG (rat) and anti-GATA6 (goat) antibodies	See Saiz et al.[22]	See Saiz et al.[22]	See Saiz et al.[22]	This paper	76
V	Saiz et al.,[22] stained with anti-NANOG (rat) and anti-GATA6 (rabbit) antibodies	See Saiz et al.[22]	See Saiz et al.[22]	See Saiz et al.[22]	This paper	27
VI	Saiz et al.,[22] stained with anti-NANOG (rabbit) and anti-GATA6 (goat) antibodies	See Saiz et al.[22]	See Saiz et al.[22]	See Saiz et al.[22]	This paper	189

allowed us to propose that expression levels of NANOG and GATA6 in the ICM during blastocyst development show a transition from local patterns in early blastocysts to a global pattern in the late blastocysts.[18] With these tools, we were further able to study how maternal age, obesity, and hyperglycemia delay cell fate decision, impacting on fertility.[19]

Traditionally, rather than studying cell fate decision based on continuous NANOG or GATA6 expression levels, different embryologists have used four distinct categories based on positive or negative expression of NANOG and GATA6.[7,8,11] Regarding positional information, the only mention was that NANOG- and GATA6-expressing cells were distributed following a salt-and-pepper pattern in mid blastocysts.[8] In this study, we aim to quantitatively characterize the salt-and-pepper pattern in the ICM that takes into account the observed 3D cell position during cell fate acquisition applying our three-dimensional neighborhood analyses. To do that, we first classify the cells into binary groups of high or low expression levels of NANOG or GATA6. We then compare the experimental spatial cell distributions with *in silico* generated artificial realizations of the common notions of the salt-and-pepper pattern using agent-based modeling. Our results reveal that in early and mid blastocysts, the binary cell fates exhibit local clustering in the ICM. The key feature of this distribution is that the cellular composition of the direct neighborhood of a cell corresponds to the cellular composition of the whole ICM.[18]

Previous approaches have implemented nearest neighbor signaling under the consideration that the FGF/MAPK pathway is the key regulator of cell fate decisions in the ICM.[20–24] Either these models did not address the spatial cell fate distributions or their results did not fit with our findings on the local patterns in early and mid blastocysts. We have recently shown that signaling range enables fine-tuning of cell patterns.[25] Based on this, we aimed to identify a potential mechanism for the generation of the cell fate distribution. Hence, we implemented two ODE modeling approaches—nearest neighbor signaling and distance-based neighbor signaling—and extended both to include cell division. Comparing the spatial cell fate distributions between the simulations and the experimental data (this study and[18] and other datasets[7,22]) reveals that cell signaling across the whole ICM best reproduces the spatial cell distribution observed experimentally in early and mid blastocysts.

RESULTS

Compiling available mouse preimplantation embryo data for independent assessment of cells expressing NANOG or GATA6

Our previous dataset for mouse blastocysts consisted of 45 embryos.[18] To obtain a more solid database that can be better integrated with data from other labs, we performed additional experiments, mainly focusing on the stages of interest, i.e., early and mid embryos. The combination of the newly acquired data with those from previous publications[7,18,22] (see STAR Methods and Table 1 for details) resulted in six datasets with 721 embryos, of which 326 were early blastocysts (32–64 cells), 218 mid blastocysts (65–90 cells), and 177 late blastocysts (more than 90 cells). The analysis of the ICM population composition using NANOG (N) and GATA6 (G6) as fate markers shows the well-known trend for N−G6−, N+G6+, N−G6−, and N−G6+ cell proportions (Figure 1A, see also Figures S1–S6 for individually analyzed datasets). In early blastocysts, most cells express both NANOG and GATA6. As development progresses, the proportion of N+G6+ cells in the ICM decreases, whereas the proportion of N+G6− cells as well as N−G6+ cells increases. The N−G6− cell proportion is not so variable. To investigate the nature of the salt-and-pepper pattern, it is essential that we classify the expression levels of NANOG and GATA6 independently (Figures 1B and 1C). First, we introduced the categories NANOG+ and NANOG− cells (Figure 1B, see STAR Methods). The assignment of the ICM cells to these categories occurred independently of their respective GATA6 levels. Our results show that the proportion of

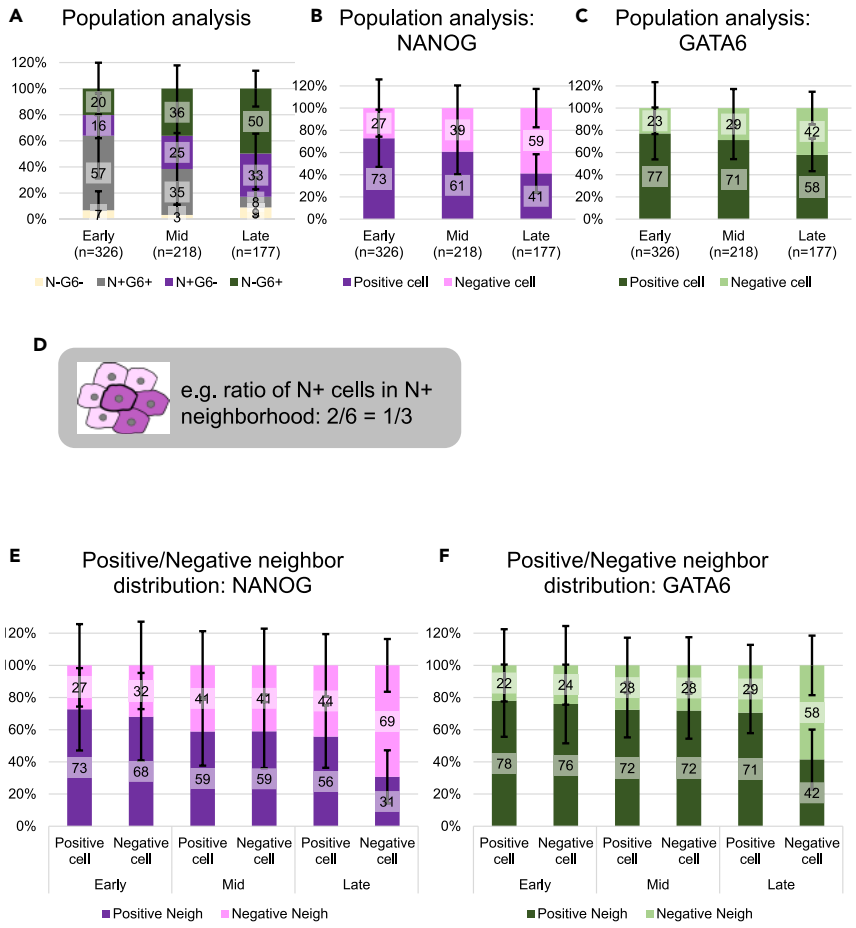


Figure 1. Average population and neighborhood composition of all embryos in all datasets

(A) Results of ICM population analysis indicating the mean percentage of N−G6−, N+G6+, N+G6− and N−G6+ cells in early (n = 326), mid (n = 218), and late (n = 177) blastocysts. Error bars show the standard deviation, here and in subsequent graphs.

Results of ICM population analysis indicating the mean percentage of N+ or N− cells (B) and G6+ or G6− cells (C) in early, mid, and late blastocysts.

(D) Neighborhood analysis description scheme: the central cell in N+ and has two N+ neighboring cells out of a total of 6 neighbors, hence 1/3 of its neighborhood is composed of N+ cells.

(E and F) Results of neighbor composition analyses for early, mid, and late blastocysts indicating the average percentage of N+ or N− neighbors (E) and G6+ or G6− neighbors (F) of each ICM cell type. See Figures S1–S6 for individually analyzed datasets.

NANOG+ cells in the ICM decreases from around 75% in early blastocysts to below 50% in late blastocysts. Categorizing the cells only according to their GATA6 levels shows a decrease in the proportion of GATA6+ cells from 75% in early to 60% in late blastocysts (Figure 1C).

Next, we analyzed the composition of the local neighborhood (Figures 1D–1F). We represented the ICM as a network and defined cell neighborhood as cell vicinity determined by the Delaunay triangulation.[26] The resulting Delaunay Cell Graph implies that neighboring cells are in direct contact.[18] The cell neighborhood composition analysis gives information about each cell type and their neighbors. Hence, we considered the mean ratio per embryo of positive neighbors of positive cells (see illustration, Figure 1D). Analogously, we included the mean ratio per embryo of negative neighbors of negative cells. For NANOG, we find that both NANOG+ and NANOG− cells have mostly NANOG+ neighbors in early and mid blastocysts. In late blastocysts, we observe a clustering such that NANOG− cells have mainly negative neighbors and NANOG+ cells have mainly positive neighbors (Figure 1E). For GATA6, we see a similar behavior (Figure 1F). In early and mid blastocysts, all cells have mostly GATA6+ neighbors, whereas in late blastocysts, GATA6+ cells cluster with GATA6+ cells and GATA6− cells cluster with GATA6− cells.

In summary, we compiled several single-cell imaging datasets of mouse embryos to increase the reliability of the analyses results. The summary statistics are characterized by a large variability between individual embryos from the same stage that remains if we consider the data from each experimental study independently (Figures S1–S6). Furthermore, we find that binary cell fates NANOG+ and GATA6+ show a clustering in late embryos, together with a decrease of the expression levels of NANOG and an increase of GATA6 with distance from the ICM centroid (Figures S1F–S6F).

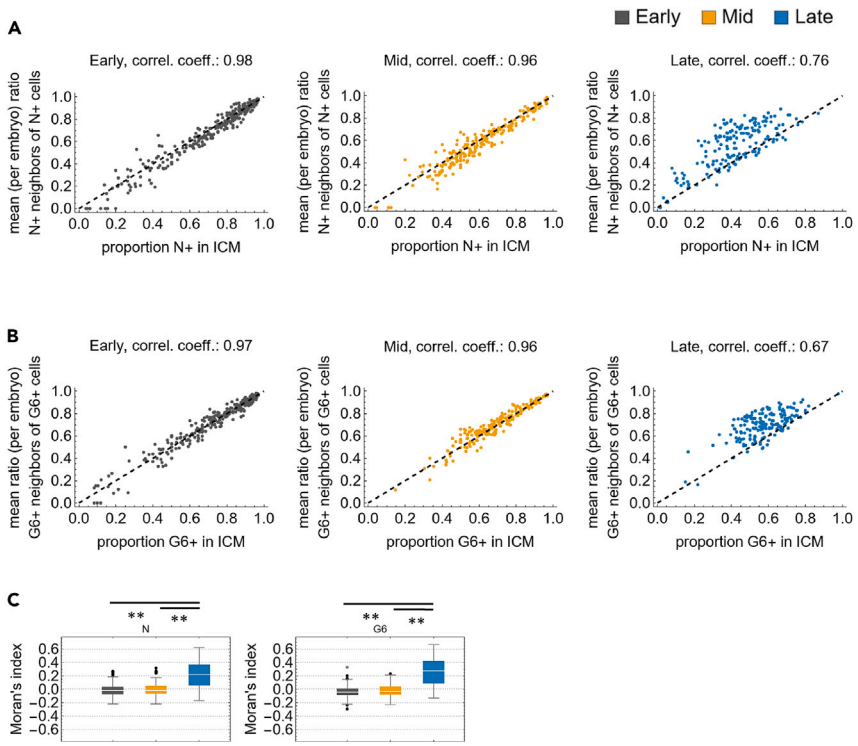


Figure 2. Single-embryo neighborhood quantification

ICM composition scatterplots showing the mean ratio of N+ neighbors of N+ cells against the proportion of N+ in each ICM (A) and the mean ratio of G6+ neighbors of G6+ cells against the proportion of G6+ in each ICM (B) in early (left, gray), mid (middle, yellow), and late (right, blue) blastocysts. Each dot represents one embryo. See Figure S7 for the plots of negative cells. Pearson's correlation coefficient is indicated above each plot.

(C) Box-Whisker-Charts of Moran's indices for the patterns of NANOG (left) and GATA6 (right) positive and negative cells for all datasets for early (gray), mid (yellow), and late (blue) blastocysts. The stars indicate significant differences at a significance level of 0.05 obtained by a Welsh's t test with Bonferroni correction.

The local cell neighborhood and global ICM population composition are strongly correlated

We have previously shown the importance of local cell neighborhood during cell fate acquisition in mouse blastocysts.[18] Others have highlighted the importance of the cell fate proportions in the ICM.[7,22] Hence, we wondered how the local cell neighborhood is related to the ICM population composition. To investigate that, we checked if there is a relation between the neighbors' composition and the overall proportion of positive and negative cells in each individual embryo (Figures 2A, 2B, and S7A–S7D). To do this, we generated ICM composition scatterplots in which we represent the mean ratio of neighbors of a specific cell type per embryo versus the total proportion of the same cell type in the whole ICM in that embryo (i.e., mean proportion of positive neighbors of a positive cell versus total positive cells in the ICM). Indeed, we find that in early and mid blastocysts, for NANOG and GATA6, for positive as well as negative cells, the average proportion of the same cell type in the direct neighborhood shows a strong correlation with the proportion of that cell type in the ICM. Furthermore, the points lie on the first bisector, indicating a proportionality with coefficient 1. Thus, if, e.g., an embryo has 20% NANOG+ cells, the NANOG+ cells will have an average of 20% NANOG+ neighbors. For late embryos, the situation changes, as the correlation decreases. The scatterplot shows a cloud structure that is denser than for early and mid embryos. Hence, the proportions of positive and negative cells in the ICM as well as the proportions of neighbors are less variable. Furthermore, these neighborhood proportions are larger than the proportion of that cell fate in the respective ICM. These results agree with the clustering of Epi and PrE cells just before the embryo implants.

Because we are interested in the spatial distribution of the different cell types in relation with their neighbors, we decided to calculate Moran's index (Figures 2C, S7E, and S7F[27]). Moran's index is a coefficient that measures the spatial autocorrelation within a dataset and has previously been applied in other biological contexts such as the spatiotemporal patterning of brain activities or cancer.[28,29] In our context of cell fate decisions, it measures how similar a cell's fate is to that of its neighbors. A value close to −1 means that the neighboring cells are of opposite type (checkerboard or period two pattern). If cell fates form large clusters, Moran's index will be positive, and it tends to 1 for a clear separation of two cell fates (sorted, see STAR Methods). We got median Moran's indices close to zero for NANOG and GATA6 in early and mid blastocysts: NANOG early (Ne): −0.01, NANOG mid (Nm): −0.001, GATA6 early (Ge): −0.05, GATA6 mid (Gm): −0.04 (Figure 2C). The Moran's index for late blastocysts is larger, consistent with the segregation of positive and negative cells.

Detailed comparisons regarding cell population composition and cell neighborhood between datasets confirmed the validity of compiling all datasets into one for further analyses (Figures S7E and S7F).

In summary, our results in late embryos show signs of separation by cell fate. For early and mid embryos, we find that the spatial cell fate distributions are stable during development. Interestingly, for each individual embryo, the local neighborhood composition is strongly correlated and comparable to the population composition of the whole ICM. Different from previous assumptions,[20] the Moran's index does not indicate a checkerboard in early and mid embryos. Now, the question arises: what kind of spatial cell fate pattern best describes these distributions.

ICM cells show characteristics of local clustering in early and mid blastocysts

To improve our understanding of the cell fate distributions in the ICM, we employed our previously established rule-based simulations of artificial patterns (Figures 3 and S8–S10[,30]). We used the three-dimensional geometry of the embryo data, i.e., the cell numbers and coordinates, and the following rules (see STAR Methods for full information about the models' implementation).

- Checkerboard (Figure 3A): positive cells have only negative neighbors.
- Alternating (Figure 3B): cell fate is alternating between nearest neighbor cells, i.e., if a cell is positive, the cell that is closest with respect to Euclidean distance is negative and so on.
- Local cluster with probability *p* (Figure 3C): one cell is randomly assigned a positive or negative state. The nearest neighbor, based on Euclidean distance, is assigned the same state as the previous cell with a given probability *p*. We used here $p = 0.55$, based on a parameter scan (Figure S11E and STAR Methods).
- Random (Figure 3D): in this case, we used the experimental cell numbers and coordinates as in the previous models, but additionally also the proportion of experimentally observed fates (Figures 1B and 1C). Then, these fate proportions were randomly distributed across the ICM of the given embryo.

The results of the three-dimensional simulated patterns of all models were plotted using the ICM composition scatterplots (Figure 3). For the checkerboard, alternating, and local cluster models, we further obtained the proportions of positive and negative cells as output or predictions (Figure S10).

The checkerboard rule generates a fate ratio of 25 $\pm$ 4% positive to 75 $\pm$ 4% negative cells (mean $\pm$ standard deviation) (Figures 3A, S8A, and S10A). According to the pattern definition, the positive cell ratio in the neighborhood of a positive cell is 0.

For the alternating model, we observe 50 $\pm$ 2% of positive and negative cells (Figures 3B, S8B, and S10B). In these two models, the results are independent of the embryo stage (Figures S8A and S8B).

Local clustering with $p = 0.55$ yields proportions of positive and negative cells in the ICM of 50 $\pm$ 11% of positive and negative cells (Figures 3C, S8C, and S10C). Looking at individual stages reveals a slight decrease in the spread of cell type proportions as development progresses (Figure S8C). The standard deviation decreases from 12% in early, to 11% in mid, and 9% in late embryos. In addition, the ratio of same cells in the direct neighborhood of a cell agrees well with the ratio of the respective fate in the whole ICM as observed experimentally (Figure 3C).

In the random model, the ratio of cells with the same fate in the direct neighborhood of a cell shows a very strong correlation with the proportion of that fate in the whole ICM simulation (Figures 3D, S9, S10D, and S10E). For very low proportions of NANOG+ or NANOG− cells in the whole ICM simulation, there are a few outliers with zero cells of the same fate in the neighborhood: these are usually due to only one or two cells of that fate that are isolated from each other in the ICM simulation. These results are again mostly independent of the embryo stage.

Comparing the ICM composition scatterplots, we observe that in the checkerboard and alternating models, the points lie mostly below the first bisector. However, in the local clustering and random models, these mostly lie on the first bisector, better representing the experimental data.

To compare the modeling results with the experimental data quantitatively, we calculated Moran's index (Figure 3E). The Moran's indices for the rule-based simulations for checkerboard, alternating, and random are all similar across stages (Figures S11A–S11D) and significantly different from the experimental data. For the Moran's indices for the local clustering model with $p = 0.55$, we could not detect a significant difference from the experimental data.

In summary, we find that the experimental data are best reproduced by a local clustering model with a clustering probability of $p = 0.55$. Hence, the question is, what kind of signaling mechanism during development could result in such a distribution.

Three-dimensional cell neighborhood analyses challenge the nearest neighbor signaling model for cell fate decisions in the ICM

Previous mathematical models of PrE differentiation rely in essence on a bi- or tristable switch in a cell combined with an intercellular signal between direct neighbors.[20–24] To test the neighborhood relations resulting from these models, we implemented a model for two hypothetical transcription factors *u* and *v* that inhibit each other in a cell and are also subject to auto-activation (Figure 4A; Tables 2 and 3).[25] In addition, we include an intercellular signal *s* that depends on the levels of u in the cell neighborhood. In the given cell, the signal has an inhibitory effect on u and an activating effect on *v*.

As a first step, we assumed that the signal is given by the average levels of *u* of the nearest neighboring cells (Figure 4A). To enable direct comparison between the simulations and the experimental data, we used ICM cell coordinates as the basis for our simulations. Furthermore, we set u−v+ as positive cell and u+v− as negative cells. Example simulations show a trend toward a checkerboard-like pattern as in,[25] although it is less clear because of the small number of cells and the three-dimensional arrangement (Figure 4B). In the ICM composition scatterplots, the ratio of same fates in the direct neighborhood lies clearly below the ratio of that fate in the whole ICM simulation for

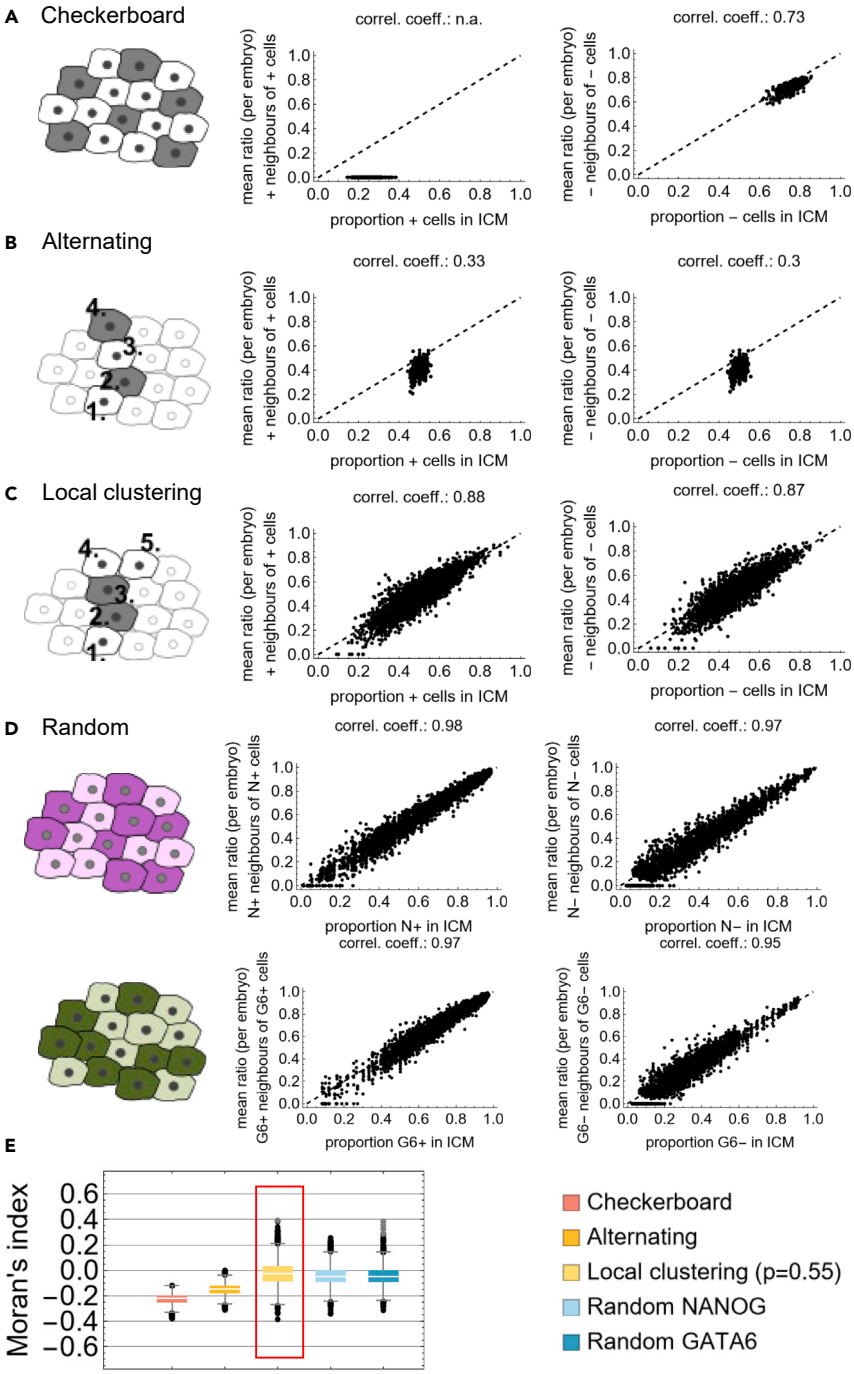


Figure 3. Single-embryo neighborhood for rule-based models

(A–D) ICM composition scatterplots for the simulation results. The first column shows a scheme for the checkerboard (A), alternating (B), local clustering (C), and random (D) models. Please note that only the illustration is in 2D, whereas the modeling is done in 3D. The second column shows the scatterplots showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells in each model ICM. The third column shows the plots for negative cells. Each dot represents one simulation for one ICM. Pearson's correlation coefficient is indicated above each plot. Please note that for the neighborhood composition of positive cells in the checkerboard pattern, Pearson's correlation coefficient is not defined. See Figures S8 and S9 for the plots separated by stage. See Figure S10 for the population analysis plots for the rule-based model patterns.

(E) Box-Whisker-Charts of Moran's indices for the rule-based model patterns checkerboard, alternating, local clustering, and random. For values separated by stage, see Figure S11. For the model results in the red box, we could not detect a significant difference to the experimental data for the four cell types N+, G6+, N−, and G6− in early and mid blastocysts. This is based on a significance level of 0.01 and Welch's t test with Bonferroni correction.

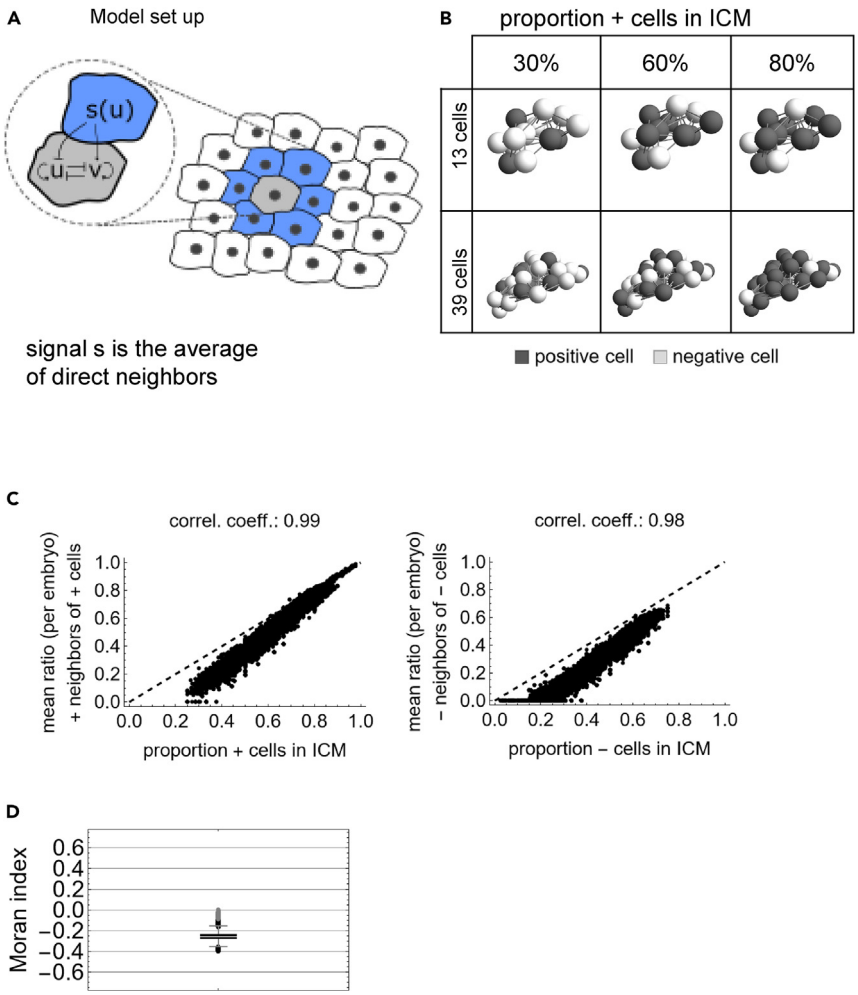


Figure 4. Nearest neighbor signaling: model setup and simulation results

(A) Illustration of the gene regulatory network considered in the nearest neighbor signaling model. Inside the cell, *u* and *v* mutually inhibit each other. In addition, *v* gets activated by an extracellular signal *s*, whereas *u* is inhibited by the same. In this model, the extracellular signal to a given cell (gray cell in the scheme) is provided by its direct neighbors (blue cells).

(B) Exemplary simulation results for three different cell fate ratios (30%, 60%, and 80% positive cells) and two cell numbers. The top represents an ICM of 13 cells corresponding to an early blastocyst. The bottom one represents an ICM of 39 cells corresponding to a late embryo.

(C) ICM composition scatterplot for the simulation results showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells (left), and the mean ratio of negative neighbors of negative cells against the proportion of negative cells in each simulated ICM (right). Each dot represents one simulation for one ICM. Pearson's correlation coefficient is indicated above each plot.

(D) Box-Whisker-Chart of Moran's indices for all simulations. See Figure S12 for results of the separation by stages.

both positive and negative cells (Figure 4C, see Figure S12A for results separated by stage). For high proportions of positive cells (>0.9), the ratio of positive cells in the neighborhood of a positive cell is identical to the proportion of positive cells in the whole ICM simulation. An interesting behavior occurs for proportions of negative cells below 0.2. Here, the mean ratio of negative neighbors of negative cells is zero. Hence, negative cells do not have negative cells in their direct neighborhood. The median Moran's index is −0.26, again hinting at a checkerboard-like pattern (Figure 4D). These results are independent of the stage, i.e., the number of cells (Figure S12B).

These results show that nearest neighbor signaling cannot recapitulate the spatial cell fate distributions in mouse blastocysts. Hence, we extended the signaling mechanism to include cells beyond the nearest neighbors.

A model in which ICM cells integrate signaling from beyond the nearest neighbors recapitulates the experimental spatial cell distribution

We assumed that cells beyond the nearest neighbors contribute to the signal for a given cell. We have previously suggested a distance-based signaling mechanism and investigated its potential for spatial pattern formation (Figure 5A;[25]). In the model, the ICM is represented by a

Table 2. List of model variables and their descriptions

Model variable	Description
u	Dimensionless concentrations of TF *U*
v	Dimensionless concentrations of TF *V*
s	Dimensionless concentrations of signal *s*

three-dimensional Delaunay cell graph.[26,31] The vertices represent the cells, and edges are drawn based on the Delaunay triangulation. This provides a very good approximation to which cells are touching.[18,26] Distances in the ICM are measured as number of edges between cells, i.e., first, second, third, … neighbor. In such a setup, the average ICM cell size does not play a role. A parameter q, with values between 0 and 1, regulates the strength of the signal dispersion in the ICM simulation (see STAR Methods and Tables 2 and 3 for more details). For small q, cells interact mainly with their direct neighbors. For $q = 0.1$, e.g., 90% of the signaling reaches the first neighbors, whereas cells further away receive only 10% of the signal. In comparison, for $q = 0.5$ the signal is halved every time it reaches cells one graph distance further away. Different values for q result in different patterns for the two cell types.[25]

We performed simulations of this model on ICM geometries that are determined by the coordinates of *in vivo* ICM cells. Example simulations on two ICMs with different cell numbers show an increase in clustering with increasing q (Figure 5B;[25]). The ICM composition scatterplots emphasize these changes in patterning with increasing q (Figure 5C see Figure S13A for results separated by stage). For positive cells and low q, the average neighborhood composition corresponds to the overall ratio of positive cells in the individual ICMs. As q increases, the curve moves above this line (Figure 5C, left). For negative cells, the results lie below the first intercept for most q. We believe the main reason for this is that our mechanism has a radial bias such that negative cells tend to accumulate at the boundary of the ICM simulation (Figure 5C, right). Moran's index supports these findings (Figures 5D and S13B). For low q, it is negative and increases with increasing q. For q above 0.8, Moran's index is around zero. This holds true if we separate the simulated data by the stages of the underlying embryo geometries (Figure S13B).

In summary, the observed spatial cell fate distribution is compatible with a model in which ICM cells integrate signaling across a longer range. In early and mid ICMs, this range can be up to five cells and in late embryos up to eight cells (Figure S13C).

Table 3. List of model parameters and initial conditions together with their values and descriptions

Model parameter	Fixed value	Description
T	108.95, 113.46, 131.74	Evolution times for early, mid, and late blastocysts
N	5,000	Number of time steps
r^*	1	Maximum radius
λ	0.083	Cell growth rate
F_0	0.01	Displacement force scaling factor
α	3	Cell stiffness
σ	0.7	Cell-cell distance optimality factor
h	0.23	Cell division distance
$-\Delta\epsilon_u$	6–7.87	Energy difference w.r.t. binding of *u*
$-\Delta\epsilon_v$	6	Energy difference w.r.t. binding of *v*
$-\Delta\epsilon_s$	2	Energy difference w.r.t. binding of *s*
$-\Delta\epsilon_{vs}$	2	Energy difference w.r.t. combined binding of *v* and *s*
r_u	1	Transcription rate of *u*
r_v	1	Transcription rate of *v*
γ_u	10	Decay rate of *u*
γ_v	10	Decay rate of *v*
q	$\in (0,1)$	Dispersion parameter for distance-based signal
u_0	$\frac{3}{4}\frac{r_u}{\gamma_u}\left(1+\frac{\xi}{10}\right)$	Initial condition for *u* and $t = 0$
v_0	$\frac{3}{4}\frac{r_v}{\gamma_v}\left(1+\frac{\xi}{10}\right)$	Initial condition for *v* and $t = 0$
ξ	$\sim N(0,1)$	Standard normal distribution noise

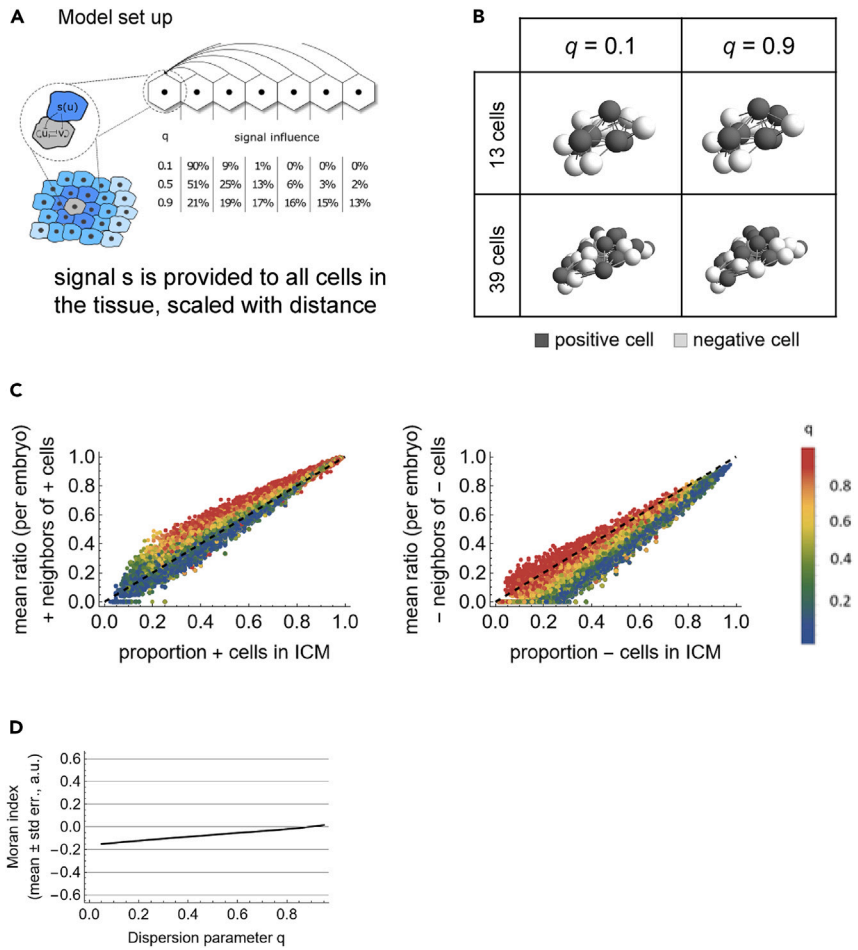


Figure 5. Distance-based neighbor signaling: model setup and simulation results

(A) Illustration of the gene regulatory network considered in the distance-based signaling model. Inside the cell, *u* and *v* mutually inhibit each other. In addition, *v* gets activated by an extracellular signal *s*, whereas *u* is inhibited by the same. In this model, the extracellular signal is the sum of all cell-cell communication between one cell and any other cell in the system. The intensity of signaling contribution depends on the dispersion parameter *q* and the distance d_{ij} between the signal-receiving cell *i* and the signal-sending cell *j*.

(B) Exemplary simulation results for two different values of the dispersion parameter *q* and two cell numbers (13—an example for an early ICM and 39—an example for a late ICM). The proportion of cells was set to 50% positive cells.

(C) ICM composition scatterplot for the simulation results showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells for each simulation for an ICM (left) and the mean ratio of negative neighbors of negative cells against the proportion of negative cells for each simulation for an ICM (right). The color code indicates the *q* values. Each dot represents one simulation for one ICM.

(D) Moran's index against dispersion parameter *q*. The plot consists of a solid line for the average and a shaded area for the standard error of the mean (SEM). Please note that the SEM is very small and therefore hardly visible.

Including cell division improves the fit of the spatial cell fate distributions for shorter-range signaling

We have previously proposed cell division as a mechanism that generates clusters in cell fate patterns.[32] Therefore, we combined the two intercellular signaling mechanisms with cell division and analyzed the resulting spatial patterns (Figure 6A). Because we did not have any specific experimental information on cell division in early and mid blastocysts, we chose as before:[32] symmetric cell division with a randomly oriented axis of division.

The ICM composition scatterplots for the model with nearest neighbor signaling that includes cell division show that the local neighborhood distribution gets closer to the proportion of the respective cells in the ICM simulation and the median Moran's index increases to −0.09 (Figures 6B, 6C, S14A, and S14B). Separating the results by the stage of the embryo geometry shows a slight increase of the Moran's index from −0.1 for early and mid to 0.06 for late blastocysts (Figure S14B). The results indicate that a model that includes nearest neighbor signaling with cell division produces a cell fate pattern that better resembles the experimental data than the model without division. However, the mean population composition does not correspond to the experimental one (Figure S14C).

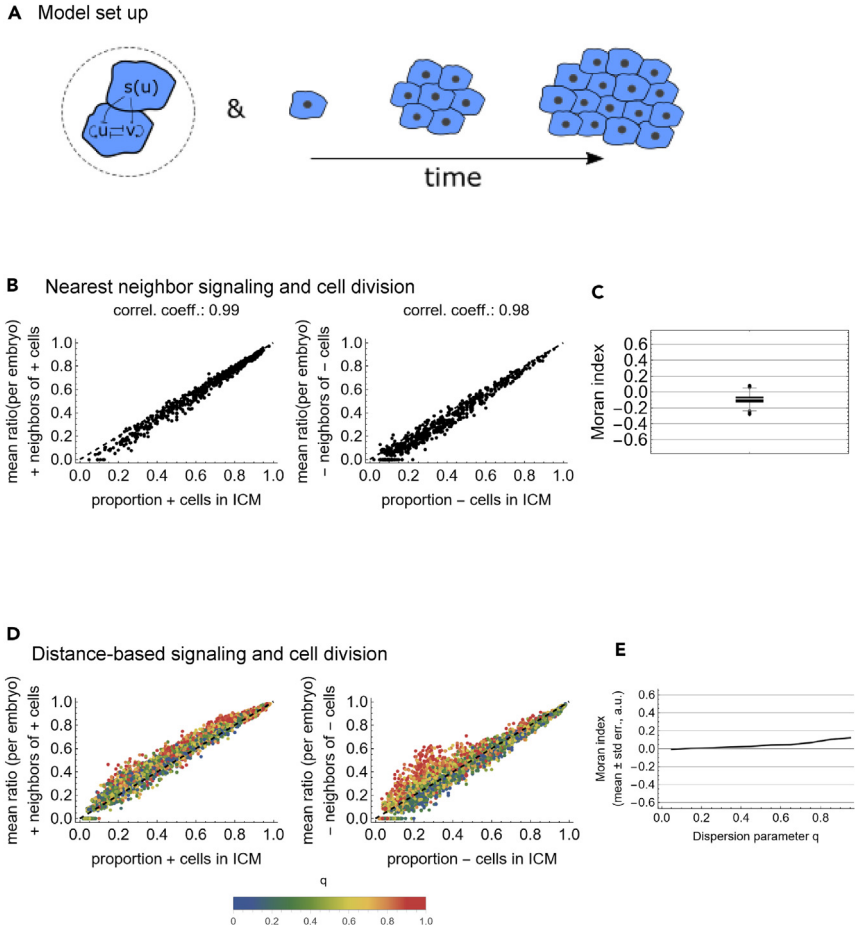


Figure 6. Nearest neighbor and distance-based neighbor signaling with cell division: model setup and simulation results

(A) Illustration of the combination of a neighbor signaling model with cell division.

(B) ICM composition scatterplot for the simulation results showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells in each simulated ICM (left) and the mean ratio of negative neighbors of negative cells against the proportion of negative cells in each simulated ICM (right). Each dot represents one simulated ICM. Pearson's correlation coefficient is indicated above each plot.

(C) Box-Whisker-Chart of Moran's indices for all simulations. For separation by stages, see Figure S14.

(D) ICM composition scatterplot for the simulation results showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells in each simulated ICM (left) and the mean ratio of negative neighbors of negative cells against the proportion of negative cells in each simulated ICM (right). The color coding indicates the *q* values. Each dot represents one simulation for one ICM.

(E) Moran's index against dispersion parameter *q*. The plot consists of a solid line for the average and a shaded area for the standard error of the mean (SEM). Please note that the SEM is very small and therefore hardly visible. For separation by stages, see Figure S14.

The inclusion of cell division in the model with distance-based signaling shifted the results such that a distribution on the first bisector with Moran's index zero is now obtained for smaller values of *q* (Figures 6D, 6E, S14D, and S14E). Separating the data by the stage of the ICM geometry also shows a stage dependence (Figures S14D and 14E). Only early and mid ICM geometries can exhibit a fate distribution on the first bisector with Moran's index zero. For late ICM geometries, the neighborhood ratios are larger than the respective cell fate proportions in the ICM and the Moran's index is larger than zero for most *q*.

In summary, including cell division in the nearest neighbor signaling model reduces the discrepancy between simulation results and experimental data. However, they still do not match the observed spatial cell fate patterns sufficiently well. Adding cell division to the distance-based model reduces the required signaling distance during cell fate decision to recapitulate the observed spatial patterns.

The spatial patterns of NANOG- and GATA6-expressing cells in early and mid blastocysts are consistent with a mechanism that integrates intercellular signaling from the whole ICM without cell division

To determine the best fitting signaling model, we compared the Moran's indices of the different models for early and mid blastocysts (Figure 7). To this end, we binned the Moran's indices for the distance-based models according to their *q* values in 10 bins of width 0.1 with bin

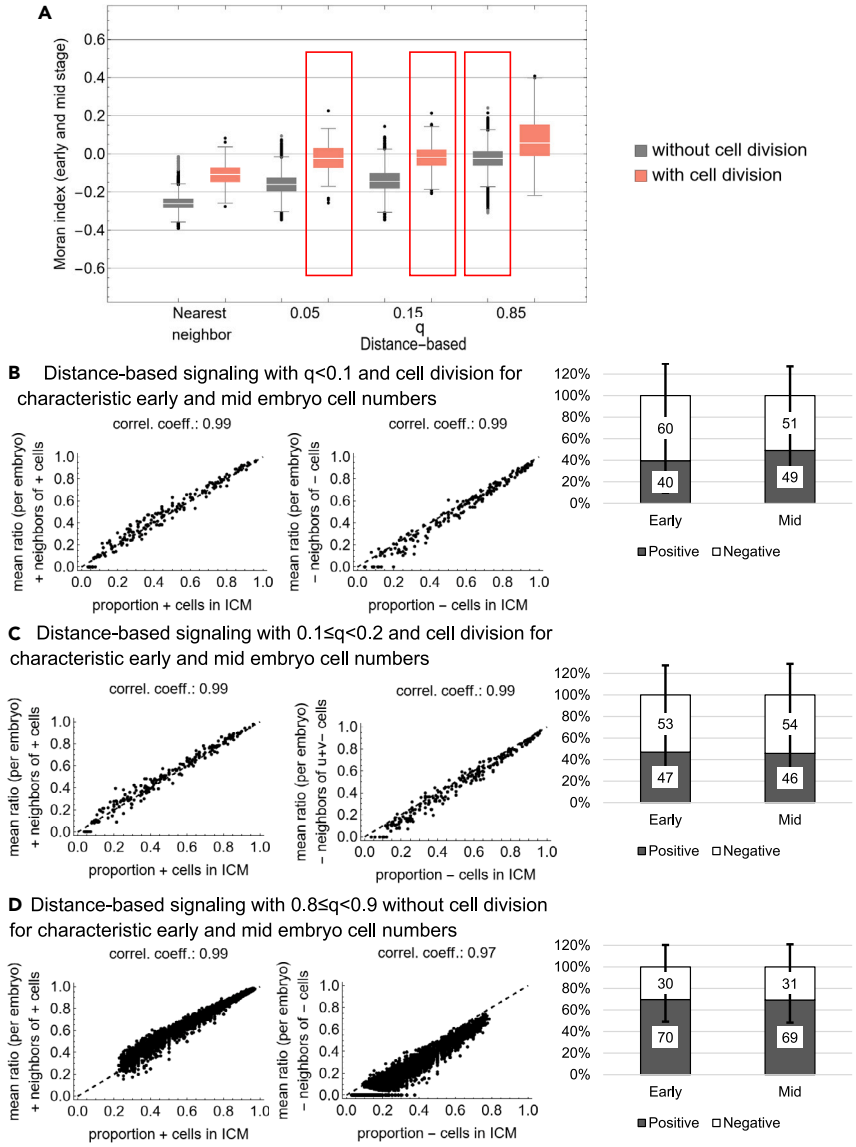


Figure 7. Best fitting signaling models

(A) Box-Whisker-Chart of Moran's indices for indicated simulations. Box-Whisker-Charts of Moran's indices for nearest neighbor and distance-based neighbor signaling without (gray) and with (light red) cell division. The distance-based neighbor signaling models were collected in bins of width 0.1 based on the respective q value. The bin centers are used as labels. For the model results in the red box, we could not detect a significant difference to the experimental data for the four cell types N+, G6+, N−, and G6− in early and mid blastocysts. We performed a Welch's t test with Bonferroni correction and a significance level of 0.05.

(B and C) ICM composition scatterplots for the simulation results for the distance-based signaling model with $q < 0.1$ (B) and $0.1 \leq q < 0.2$ (C) and cell division. Scatterplots showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells for each simulation for an ICM (left) and the mean ratio of negative neighbors of negative cells against the proportion of negative cells for each simulation for an ICM (center). Each dot represents one simulation for one ICM of early and mid embryos. (Right) ICM population analyses indicating the average percentage of positive and negative cells in the simulations for early, mid, and late ICMs.

(D) ICM composition scatterplots for the simulation results for the distance-based signaling model with $0.8 \leq q < 0.9$ and no cell division. Scatterplots showing the mean ratio of positive neighbors of positive cells against the proportion of positive cells for each simulation for an ICM (left) and the mean ratio of negative neighbors of negative cells against the proportion of negative cells for each simulation for an ICM (center). Each dot represents one simulation for one ICM of early and mid embryos. (Right) ICM population analysis indicating the average percentage of positive and negative cells in the simulations for early, mid, and late ICMs.

centers at 0.05, 0.15, etc. We found three models for which we could not detect a statistically significant difference to the experimental data for NANOG and GATA6 patterns in early and mid embryos: distance-based signaling with $q = 0.05$ and $q = 0.15$ and cell division as well as distance-based signaling with $q = 0.85$ and no cell division. For the ICM composition scatterplots, we see that both models show a strong correlation between the local neighborhood composition and the ICM composition (Figures 5B, 5C, and S15). Furthermore, the points lie on the first bisector.

Looking at the mean cell fate ratios however provides a clear distinction (Figures 7B left and 7C left). The model with cell division generates on average between 60%–50% positive and 40%–50% negative cells and no neighbor clustering effect of any cell type (Figures S15B and S15C). However, the model without cell division results in 70% positive and 30% negative cells (Figure 7D, left) and clustering of positive cells (Figure S15D). The latter is closer to the observed experimental data (Figure 1).

Overall, we conclude that a signaling mechanism that integrates the signals from all ICM cells with no cell division can yield the local cell fate clusters in mouse ICMs and additionally reproduce the mean population proportions observed in the ICM.

DISCUSSION

It has been proposed that during cell fate acquisition in mouse preimplantation embryos, Epi and PrE progenitors are located following a salt-and-pepper pattern.[8,9] Traditionally, this pattern has been assumed to be random[8–10] or checkerboard-like[20] without investigating its true nature. Rather, cell fate acquisition has been mainly studied focusing on cell fate ratios, with FGF/MAPK signaling as the main pathway involved in the process.[21] However, other signaling pathways are also involved, modulating the cell fate acquisition process.[33,34] Very few studies have considered the spatial cell fate distribution in their signaling models for the embryo[20,21] and always without quantitative experimental data for comparison. We have previously used single-cell immunofluorescence imaging data of early, mid, and late blastocysts; neighborhood analyses; and rule-based modeling to analyze the patterns in NANOG and GATA6 expression levels.[18] We have found that Epi progenitor cells (NANOG+/GATA6−) form clusters in early and mid blastocysts and that NANOG expression correlates with the number of neighboring cells. In addition, GATA6 expression correlates with the expression of neighboring cells.[18] In this study, we developed a complementary approach. We considered NANOG and GATA6 separately and instead of analyzing continuous expression levels, we focused on binary cell states: NANOG positive or negative as well as GATA6 positive or negative. This approach has several advantages: (1) we can merge experimental data from different microscopes or immunofluorescent experiments that typically vary in fluorescence intensity, as well as (2) generated by different labs; (3) we can relate our results to the salt-and-pepper idea as this implicitly assumes binary cell fates; and (4) we can link our results to the output of our simulations based on ordinary differential equations (ODEs) for a toggle switch coupled to intercellular signaling. This last step is important to gain a mechanistic understanding that we could not obtain from the expression level data. We show that the salt-and-pepper pattern in early and mid mouse ICMs is best represented by a local clustering pattern resulting from a signaling mechanism with a large dispersion range and no cell division. This result agrees with our previous findings of correlations for GATA6 expression levels between neighbors and the clusters of Epi progenitor cells. Furthermore, it fits with previous findings in two- and three-dimensional *in vitro* cell cultures by others and us, respectively.[30,35] Raina et al. observed a clustering effect in 2D in *in vitro* cell cultures, which they could fit well by a model that considers nearest and second nearest neighbors for the signaling mechanism.[35] Our previous work with ICM organoids has revealed that the salt-and-pepper pattern of the *in vitro* system fits a local clustering pattern[30] and that predicting the fate of a given cell based on the neighboring fates using neural networks only shows a 70% accuracy.[36]

Previous ODE models for cell fate decision in early mouse embryos have incorporated intracellular inhibition and intercellular signaling only between nearest neighbors in combination with cell division.[20,22] Tosenberger et al. analyzed the three-dimensional cell fate distribution of the simulations but have not presented experimental data on this part.[20] Saiz et al. focused on cell fate ratios in late embryos. Cell fate ratios or the spatial distribution during differentiation (early and mid embryos) were not considered.[22] We implemented an analogous nearest neighbor signaling model with cell division and found that the simulation results for the spatial cell distribution were not comparable to the experimental data. Interestingly, our results further suggest that in early and mid embryos, it is not the absolute cell fate ratio that is conserved. Instead, the relation between the cell fate ratios in the neighborhood and in the whole ICM stays constant.

In our study, the model showing the best fit with the large experimental data integrates signaling from nearest neighbors as well as from cells further away in the ICM, which have a considerable influence. Our analyses show that in early and mid ICMs, this can range to up to the fifth neighbor.

The main difference between nearest neighbor signaling and distance-based signaling is how far the signal travels. The main signal in this context is FGF4,[13] and it is usually assumed that it freely diffuses within the ICM. On a timescale of several days, FGF4 has been shown to diffuse over many millimetres in cell culture dishes and that embryonic stem cells use it for communication across these large scales.[37] For a timescale of 12 h, FGF4 signaling reaches mainly first and second neighbors in embryonic stem cell cultures.[35] FGF/MAPK signaling range can be modulated by several mechanisms: FGF diffusion via heparan sulfate proteoglycans at the extracellular matrix, FGFR trafficking, and *fgf4* mRNA decay.[38] All these cellular mechanisms have yet to be investigated in the developing mouse preimplantation embryo. Moreover, FGFs might not even diffuse at all, as has been shown for other morphogens like Wingless.[39] In this context, it might also be possible that FGF/MAPK signaling range is increased by cell cytonemes present in the ICM cells[40] or microvesicles.[41] Furthermore, not only FGF4 is expressed in the early mouse embryo, other FGFs like FGF10, FGF3, and FGF5 are also expressed, complicating the FGF/MAPK signaling scenario.[42] Another cellular process that might have an effect in the cell neighborhood and signal dispersion is the presence of cytoplasmic bridges

connecting sister cells during embryonic development hence coordinating cell fate decisions.[43] These bridges have been observed in mouse morulae,[44,45] as well as in Epi cells from postimplantation embryos.[46,47]

The best fitting model in our study does not incorporate cell division. This indicates that cell division is not absolutely required to generate the cell fate patterns observed in early and mid blastocysts. Furthermore, we see that it even hinders obtaining the experimentally observed mean population distribution. This suggests that the assumptions for cell division implemented in the model might not be adequate. Cell division is currently implemented symmetrically and with a random division axis because previous results for modeling the neighborhood composition of ICM organoids have shown that symmetric division fits the experimental distributions best.[32] In that regard, mouse embryonic stem cells expressing a *Nanog*-GPF reporter divide symmetrically during differentiation.[48] However, an asymmetric distribution of the two transcription factors to the daughter cells could be plausible and might account for the differences in population distribution. Because the possibilities for implementing cell division mechanisms are huge, new experimental data to narrow down potential cell division mechanisms are required for future model refinements.

Furthermore, a factor that we have neglected so far, is apoptosis. On average, 3.2 ± 2.81 dead cells occur in the mouse ICM.[49] Apoptosis mostly starts occurring in mid blastocysts, and the rates steadily increase with increasing cell number.[8,50] Hence, apoptosis could only account for the discrepancy for mid blastocysts. Furthermore, our model would predict that apoptosis of one cell fate dominates over the other. Our model with cell division currently predicts equal amounts of positive and negative cells. The experimental data show a distribution of on average 30% negative and 70% positive cells both for NANOG and GATA6. To match this, an increase in cell death in negative cells would be required. Overall, to further investigate the role of cell division in cell differentiation patterning, a signaling model in which differential apoptosis and asymmetric cell division is implemented would be required.

Finally, our model and local clustering pattern during cell fate acquisition in mouse preimplantation embryo is only applicable to early and mid blastocysts and does not fit late ones. This is likely due to not only enhanced cell division asymmetries or apoptosis but also cell migration of PrE cells to their final position facing the cavity before the embryo implants. Another factor influencing cell fate distribution in late blastocysts might be the FGF/MAPK signaling maturation, with changes in the expression levels of its components and activity.[14,42] These components need to be considered when expanding our model to late blastocysts.

In summary, we are proposing a novel mechanism that better aligns with previously overlooked spatial cell fate patterning compared with existing models. Our results reveal unknown questions in the field such as the exact details of cell division that need to be investigated with further experiments. Consequently, our work represents a significant stride toward enhancing our comprehension of Epi versus PrE differentiation, although it is important to note that this marks just one chapter in the ongoing story.

It has been proposed that rabbit embryos also exhibit a salt-and-pepper pattern.[51] Since our approach, including the patterning in our ODE model, is independent of cell number, it could be directly applied to these larger embryos. The notion of a salt-and-pepper pattern has also come up in other developmental questions like the cell fate segregation in pancreatic or neural development through Notch signaling,[52,53] the segregation of the definitive endoderm,[54] mammalian visual cortex organization,[55] or the cilia in the notochord.[56] Focusing on the spatial statistics in the data analysis and the model fitting might also help to reveal the nature of the spatial pattern and improve the understanding of the underlying mechanisms in those contexts. In summary, using spatial statistics on single-cell embryo data reveals that understanding cell fate decisions in mouse blastocysts must accommodate not only a cells neighborhood but also a larger proportion of the ICM.[26] This ensures that cells make coordinated decisions to successfully make an embryo.

Limitations of the study

The experimental data used in this study come from immunostained fixed embryos. Ideally, the experimental data should have been obtained from time-lapse imaging data from double reporter mouse lines reporting NANOG and GATA6 protein levels. To our knowledge such a mouse line is currently not available.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental data
  - Modelling and simulations
- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108106.

ACKNOWLEDGMENTS

Work at SMD lab was funded by the ACIISI (CEI2019-02), Programa de Ayudas a la Investigación de la ULPGC, and ACIISI co-funded by FEDER Funds (ProID2020010013). JLG is supported by the ULPGC predoctoral program. SMD was supported by the "Viera y Clavijo" Program from the Agencia Canaria de Investigación, Innovación y Sociedad de la Información (ACIISI), and the ULPGC. Imaging was done at the SIMACE (Servicio de Investigación en Microscopía Avanzada Confocal y Electrónica), at the IUIBS. Work at the SCF lab was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 470129398 and start-up funding by the University of Wuerzburg. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

AUTHOR CONTRIBUTIONS

Conceptualization: S.C.F. and S.M.D.; Methodology: S.C.F., S.S., and S.M.D.; Validation: S.C.F. and S.M.D.; Formal Analysis: S.C.F., S.S., and S.M.D.; Investigation: J.L.G. and S.M.D.; Resources: S.C.F. and S.M.D.; Data Curation: S.C.F. and S.M.D.; Writing—Original Draft: S.C.F. and S.M.D., Writing—Review & Editing: S.C.F., S.S., J.L.G., and S.M.D.; Visualization: S.C.F. and S.M.D.; Supervision: S.C.F. and S.M.D.; Project Administration: S.C.F. and S.M.D.; Funding Acquisition: S.C.F., J.L.G., and S.M.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: May 8, 2023
Revised: August 3, 2023
Accepted: September 26, 2023
Published: September 30, 2023

REFERENCES

1. Briscoe, J. (2019). Understanding Pattern Formation in Embryos: Experiment, Theory, and Simulation. J. Comput. Biol. *26*, 696–702. https://doi.org/10.1089/CMB.2019.0090.
2. Simsek, M.F., and Özbudak, E.M. (2022). Patterning principles of morphogen gradients. Open Biol. *12*, 220224. https://doi.org/10.1098/RSOB.220224.
3. Marshall, F.H.A. (1911). The Works of Aristotle - The Works of Aristotle. Translated into English under the Editorship of J. A. Smith and W. D. Ross. Vol. IV. Historia Animalium, by D'Arcy Wentworth Thompson. Oxford: Clarendon Press, 1910. Price 10s. 6d. net. Classical Rev. *25*, 208–209. https://doi.org/10.1017/S0009840X00047223.
4. Płusa, B., and Piliszek, A. (2020). Common Principles of Early Mammalian Embryo Self-Organisation. Development *147*, dev183079. https://doi.org/10.1242/DEV.183079.
5. Tsichlaki, E., and Fitzharris, G. (2016). Nucleus downscaling in mouse embryos is regulated by cooperative developmental and geometric programs. Sci. Rep. *6*, 28040. https://doi.org/10.1038/SREP28040.
6. Forsyth, J.E., Al-Anbaki, A.H., de la Fuente, R., Modare, N., Perez-Cortes, D., Rivera, I., Seaton Kelly, R., Cotter, S., and Plusa, B. (2021). IVEN: A quantitative tool to describe 3D cell position and neighbourhood reveals architectural changes in FGF4-treated preimplantation embryos. PLoS Biol. *19*, e3001345. https://doi.org/10.1371/JOURNAL.PBIO.3001345.
7. Saiz, N., Williams, K.M., Seshan, V.E., and Hadjantonakis, A.-K. (2016). Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. Nat. Commun. *7*, 13463. https://doi.org/10.1038/ncomms13463.
8. Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development *135*, 3081–3091. https://doi.org/10.1242/dev.021519.
9. Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. Dev. Cell *10*, 615–624. https://doi.org/10.1016/j.devcel.2006.02.020.
10. Rossant, J., Chazaud, C., Yamanaka, Y., Jones, M., Robertson, E.J., Smith, A., and Edwards, R.G. (2003). Lineage allocation and asymmetries in the early mouse embryo. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 1341–1348. , discussion 1349. https://doi.org/10.1098/RSTB.2003.1329.
11. Frankenberg, S., Gerbe, F., Bessonnard, S., Belville, C., Pouchin, P., Bardot, O., and Chazaud, C. (2011). Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. Dev. Cell *21*, 1005–1013. https://doi.org/10.1016/j.devcel.2011.10.019.
12. Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. Dev. Cell *18*, 675–685. https://doi.org/10.1016/j.devcel.2010.02.012.
13. Kang, M., Piliszek, A., Artus, J., and Hadjantonakis, A.-K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development *140*, 267–279. https://doi.org/10.1242/dev.084996.
14. Kang, M., Garg, V., and Hadjantonakis, A.K. (2017). Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. Dev. Cell *41*, 496–510.e5. https://doi.org/10.1016/j.devcel.2017.05.003.
15. Messerschmidt, D.M., and Kemler, R. (2010). Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. Dev. Biol. *344*, 129–137. https://doi.org/10.1016/J.YDBIO.2010.04.020.
16. Molotkov, A., and Soriano, P. (2018). Distinct mechanisms for PDGF and FGF signaling in primitive endoderm development. Dev. Biol. *442*, 155–161. https://doi.org/10.1016/j.ydbio.2018.07.010.
17. Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., and Hiiragi, T. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat. Cell Biol. *16*, 27–37. https://doi.org/10.1038/ncb2881.
18. Fischer, S.C., Corujo-Simon, E., Lilao-Garzon, J., Stelzer, E.H.K., and Muñoz-Descalzo, S. (2020). The transition from local to global patterns governs the differentiation of mouse blastocysts. PLoS One *15*, e0233030. https://doi.org/10.1371/journal.pone.0233030.
19. Lilao-Garzón, J., Brito-Casillas, Y., Quesada-Canales, O., Wägner, A.M., and Munoz-Descalzo, S. (2023). Maternal age, obesity and hyperglycaemia are associated with a delay in preimplantation embryo development in mouse. Reproduction *1*. https://doi.org/10.1530/REP-23-0024.
20. Tosenberger, A., Gonze, D., Bessonnard, S., Cohen-Tannoudji, M., Chazaud, C., and

Dupont, G. (2017). A multiscale model of early cell lineage specification including cell division. NPJ Syst. Biol. Appl. *3*, 16. https://doi.org/10.1038/s41540-017-0017-0.

21. Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. Development *141*, 3637–3648. https://doi.org/10.1242/dev.109678.
22. Saiz, N., Mora-Bitria, L., Rahman, S., George, H., Herder, J.P., Garcia-Ojalvo, J., and Hadjantonakis, A.-K. (2020). Growth factor-mediated coupling between lineage size and cell fate choice underlies robustness of mammalian development. Elife *9*, e56079. https://doi.org/10.7554/eLife.56079.
23. Schröter, C., Rué, P., Mackenzie, J.P., and Martinez Arias, A. (2015). FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. Development *142*, 4205–4216. https://doi.org/10.1242/dev.127530.
24. Stanoev, A., Schröter, C., and Koseska, A. (2021). Robustness and Timing of Cellular Differentiation through Population-Based Symmetry Breaking. Development *148*, dev197608. https://doi.org/10.1242/DEV.197608.
25. Schardt, S., Fischer, S.C., Schardt, S., and Fischer, S.C. (2023). Adjusting the range of cell–cell communication enables fine-tuning of cell fate patterns from checkerboard to engulfing. J. Math. Biol. *87*, 1–29. https://doi.org/10.1007/S00285-023-01959-9.
26. Fischer, S.C., Bassel, G.W., and Kollmannsberger, P. (2023). Tissues as networks of cells: towards generative rules of complex organ development. J. R. Soc. Interface *20*, 20230115. https://doi.org/10.1098/RSIF.2023.0115.
27. Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika *37*, 17–23. https://doi.org/10.2307/2332142.
28. Braz Junior, G., Cardoso de Paiva, A., Corrêa Silva, A., and Cesar Muniz de Oliveira, A. (2009). Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM. Comput. Biol. Med. *39*, 1063–1072. https://doi.org/10.1016/J.COMPBIOMED.2009.08.009.
29. Schmal, C., Myung, J., Herzel, H., and Bordyugov, G. Moran's I quantifies spatio-temporal pattern formation in neural imaging data. Bioinformatics. 33, 3072-3079. 10.1093/bioinformatics/btx351
30. Mathew, B., Muñoz-Descalzo, S., Corujo-Simon, E., Schröter, C., Stelzer, E.H.K., and Fischer, S.C. (2019). Mouse ICM Organoids Reveal Three-Dimensional Cell Fate Clustering. Biophys. J. *116*, 127–141. https://doi.org/10.1016/J.BPJ.2018.11.011.
31. Schmitz, A., Fischer, S.C., Mattheyer, C., Pampaloni, F., and Stelzer, E.H.K. (2017). Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids. Sci. Rep. *7*, 43693. https://doi.org/10.1038/srep43693.
32. Liebisch, T., Drusko, A., Mathew, B., Stelzer, E.H.K., Fischer, S.C., and Matthäus, F. (2020). Cell fate clusters in ICM organoids arise from cell fate heredity and division: a modelling approach. Sci. Rep. *10*, 22405. https://doi.org/10.1038/s41598-020-80141-3.
33. Thamodaran, V., and Bruce, A.W. (2016). p38 (Mapk14/11) occupies a regulatory node governing entry into primitive endoderm differentiation during preimplantation mouse embryo development. Open Biol. *6*, 160190. https://doi.org/10.1098/RSOB.160190.
34. Goissis, M.D., Bradshaw, B., Posfai, E., and Rossant, J. (2023). Influence of FGF4 and BMP4 on FGFR2 dynamics during the segregation of epiblast and primitive endoderm cells in the pre-implantation mouse embryo. PLoS One *18*, e0279515. https://doi.org/10.1371/JOURNAL.PONE.0279515.
35. Raina, D., Bahadori, A., Stanoev, A., Protzek, M., Koseska, A., and Schroter, C. (2021). Cell-cell Communication through FGF4 Generates and Maintains Robust Proportions of Differentiated Cell Types in Embryonic Stem Cells. Development *148*, dev199926. https://doi.org/10.1242/DEV.199926.
36. Dirk, R., Fischer, J.L., Schardt, S., Ankenbrand, M.J., and Fischer, S.C. (2022). Recognition and Reconstruction of Cell Differentiation Patterns with Deep Learning. https://doi.org/10.48550/arxiv.2212.10058.
37. Daneshpour, H., van den Bersselaar, P., Chao, C.-H., Fazzio, T.G., and Youk, H. (2023). Macroscopic quorum sensing sustains differentiating embryonic stem cells. Nat. Chem. Biol. *19*, 596–606. https://doi.org/10.1038/s41589-022-01225-x.
38. Balasubramanian, R., and Zhang, X. (2016). Mechanisms of FGF gradient formation during embryogenesis. Semin. Cell Dev. Biol. *53*, 94–100. https://doi.org/10.1016/j.semcdb.2015.10.004.
39. Alexandre, C., Baena-Lopez, A., and Vincent, J.P. (2013). Patterning and growth control by membrane-tethered Wingless. Nature *505*, 180–185. https://doi.org/10.1038/nature12879.
40. Salas-Vidal, E., and Lomelí, H. (2004). Imaging filopodia dynamics in the mouse blastocyst. Dev. Biol. *265*, 75–89. https://doi.org/10.1016/j.ydbio.2003.09.012.
41. Desrochers, L.M., Bordeleau, F., Reinhart-King, C.A., Cerione, R.A., and Antonyak, M.A. (2016). Microvesicles provide a mechanism for intercellular communication by embryonic stem cells during embryo implantation. Nat. Commun. *7*, 11958. https://doi.org/10.1038/NCOMMS11958.
42. Nowotschin, S., Setty, M., Kuo, Y.Y., Liu, V., Garg, V., Sharma, R., Simon, C.S., Saiz, N., Gardner, R., Boutet, S.C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. Nature *569*, 361–367. https://doi.org/10.1038/s41586-019-1127-1.
43. Hawdon, A., Aberkane, A., and Zenker, J. (2021). Microtubule-dependent Subcellular Organisation of Pluripotent Cells. Development *148*, dev199909. https://doi.org/10.1242/DEV.199909/272646.
44. Pomp, O., Lim, H.Y.G., Skory, R.M., Moverley, A.A., Tetlak, P., Bissiere, S., and Plachta, N. (2022). A monoastral mitotic spindle determines lineage fate and position in the mouse embryo. Nat. Cell Biol. *24*, 155–167. https://doi.org/10.1038/s41556-021-00826-3.
45. Sołtyńska, M.S. (1982). The possible mechanism of cell positioning in mouse morulae: an ultrastructural study. Development *68*, 137–147. https://doi.org/10.1242/DEV.68.1.137.
46. Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development *113*, 891–911. https://doi.org/10.1242/DEV.113.3.891.
47. Gardner, R.L., and Cockroft, D.L. (1998). Complete dissipation of coherent clonal growth occurs before gastrulation in mouse epiblast. Development *125*, 2397–2402. https://doi.org/10.1242/DEV.125.13.2397.
48. Nakamura, S., Maruyama, A., Kondo, Y., Kano, A., De Sousa, O.M., Iwahashi, M., Hexig, B., Akaike, T., Li, J., Hayashi, Y., and Ohnuma, K. (2018). Asymmetricity Between Sister Cells of Pluripotent Stem Cells at the Onset of Differentiation. Stem Cell. Dev. *27*, 347–354. https://doi.org/10.1089/scd.2017.0113.
49. Pisko, J., Špirková, A., Čikoš, Š., Olexiková, L., Kovaříková, V., Šefčíková, Z., and Fabian, D. (2021). Apoptotic cells in mouse blastocysts are eliminated by neighbouring blastomeres. Sci. Rep. *11*, 9228. https://doi.org/10.1038/S41598-021-88752-0.
50. Soszyńska, A., Filimonow, K., Wigger, M., Wołukanis, K., Gross, A., Szczepańska, K.B., and Suwińska, A. (2023). Multi-level Fgf4- and apoptosis-dependent regulatory mechanism ensures the plasticity of ESC-chimaeric mouse embryo. Development *150*, dev201756. https://doi.org/10.1242/DEV.201756.
51. Piliszek, A., Madeja, Z.E., and Plusa, B. (2017). Suppression of ERK signalling abolishes primitive endoderm formation but does not promote pluripotency in rabbit embryo. Development (Camb.) *144*, 3719–3730. https://doi.org/10.1242/dev.156406.
52. Kageyama, R., Ohtsuka, T., Shimojo, H., and Imayoshi, I. (2008). Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. Nat. Neurosci. *11*, 1247–1251. https://doi.org/10.1038/nn.2208.
53. Xu, X., Seymour, P.A., Sneppen, K., Trusina, A., La, A., Egeskov-Madsen, R., Jørgensen, M.C., Jensen, M.H., and Serup, P. Jag1-Notch Cis-Interaction Determines Cell Fate Segregation in Pancreatic Development. 10.1038/s41467-023-35963-w
54. Pour, M., Kumar, A.S., Farag, N., Bolondi, A., Kretzmer, H., Walther, M., Wittler, L., Meissner, A., and Nachman, I. (2022). Emergence and patterning dynamics of mouse-definitive endoderm. iScience *25*, 103556. https://doi.org/10.1016/J.ISCI.2021.103556.
55. Jang, J., Song, M., and Paik, S.B. (2020). Retino-Cortical Mapping Ratio Predicts Columnar and Salt-and-Pepper Organization in Mammalian Visual Cortex. Cell Rep. *30*, 3270–3279.e3. https://doi.org/10.1016/J.CELREP.2020.02.038.
56. Feistel, K., and Blum, M. (2006). Three types of cilia including a novel 9+4 axoneme on the notochordal plate of the rabbit embryo. Dev. Dynam. *235*, 3348–3358. https://doi.org/10.1002/DVDY.20986.
57. Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215–3222. https://doi.org/10.1242/dev.038893.
58. Lou, X., Kang, M., Xenopoulos, P., Muñoz-Descalzo, S., and Hadjantonakis, A.K. (2014). A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data. Stem Cell Rep. *2*, 382–397. https://doi.org/10.1016/J.STEMCR.2014.01.010.
59. Hammer, D.A.T., Ryan, P.D., Hammer, Ø., and Harper, D.A.T. (2001). Past: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontol. Electron. *4*, 178.

60. Lloyd, A.C. (2013). The Regulation of Cell Size. Cell *154*, 1194–1205. https://doi.org/10.1016/J.CELL.2013.08.053.

61. Conlon, I., and Raff, M. (2003). Differences in the way a mammalian cell and yeast cells coordinate cell growth and cell-cycle progression. J. Biol. *2*, 7–10. https://doi.org/10.1186/1475-4924-2-7.

62. Mir, M., Wang, Z., Shen, Z., Bednarz, M., Bashir, R., Golding, I., Prasanth, S.G., and Popescu, G. (2011). Optical measurement of cycle-dependent cell growth. Proc. Natl. Acad. Sci. USA *108*, 13124–13129. https://doi.org/10.1073/PNAS.1100506108.

63. Dolznig, H., Grebien, F., Sauer, T., Beug, H., and Müllner, E.W. (2004). Evidence for a size-sensing mechanism in animal cells. Nat. Cell Biol. *6*, 899–905. https://doi.org/10.1038/ncb1166.

64. Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G., and Kirschner, M.W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature *494*, 480–483. https://doi.org/10.1038/nature11897.

65. Lin, J., and Amir, A. (2020). From single-cell variability to population growth. Phys. Rev. E *101*, 012401. https://doi.org/10.1103/PHYSREVE.101.012401.

66. Stichel, D., Middleton, A.M., Müller, B.F., Depner, S., Klingmüller, U., Breuhahn, K., and Matthäus, F. (2017). An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. NPJ Syst. Biol. Appl. *3*, 1–10. https://doi.org/10.1038/s41540-017-0006-3.

67. Middleton, A.M., Fleck, C., and Grima, R. (2014). A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. J. Theor. Biol. *359*, 220–232. https://doi.org/10.1016/J.JTBI.2014.06.011.

68. Van Rossum, G., and Drake, F.L. (2009). Python 3 Reference Manual (CreateSpace).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-NANOG	ThermoFisher Scientific	Catalog # 14-5761-80; RRID:AB_763613
Goat polyclonal anti-GATA6	R&D Systems	Cat# AF1700; RRID:AB_2108901
Donkey Alexa Fluor 488 – conjugated anti-Rat	Invitrogen	Catalog # A-21208
Donkey Alexa Fluor 568 – conjugated anti-Goat	Invitrogen	Catalog # A-11057
Chemicals, peptides, and recombinant proteins		
DAPI	ThermoFisher Scientific (Invitrogen)	D1306
M2 (Embryomax)	Merck (Millipore)	MR-015-D
BSA	Merck (Sigma)	10735086001
PVP	Merck (Sigma)	P0930
Triton X-100	Merck (Sigma)	X-100
Tween	Merck (Sigma)	P1379
Donkey Serum	Merck (Sigma)	D9663
Vectashield	Vector Labs	H-1200
PBS	Capricorn Scientific	PBS-1A
PFA	Merck (Sigma)	P6148
Acid Tyrodes	Merck (Sigma)	T1788-100ML
Deposited data		
Data set I: New embryo raw data and complete analysis	This paper	https://doi.org/10.5281/zenodo.7867284
Data set II: Embryo data raw data, image analysis, population analysis	Fischer et al.[18]	https://zenodo.org/record/7486867
Data set II: Neighbourhood extraction	This paper	https://doi.org/10.5281/zenodo.7867284
Data set III: Embryo data raw data, image analysis, population analysis	Saiz et al.[7]	https://doi.org/10.6084/m9.figshare.c.3447537.v1
Data set III: Neighbourhood extraction	This paper	https://doi.org/10.5281/zenodo.7867284
Data set IV-VI: Embryo data raw data, image analysis, population analysis	Saiz et al.[22]	https://doi.org/10.6084/m9.figshare.c.4736507
Data set IV-VI: Neighbourhood extraction	This paper	https://doi.org/10.5281/zenodo.7867284
Code used to transform and analyse data	This paper	https://doi.org/10.5281/zenodo.7867284
Experimental models: Organisms/strains		
Mouse: Swiss	Bred in-house	
Software and algorithms		
MINS	Lou et al.[58]	https://github.com/therealkatlab/MINS
Matlab	MathWorks	https://www.mathworks.com/products/matlab.html
Mathematica	Wolfram	https://www.wolfram.com/mathematica/
ZEN	Carl Zeiss Microsystems	https://www.zeiss.com/microscopy/us/products/microscope-software/zen-lite.html
ImageJ	NIH	https://imagej.nih.gov/ij/
PAST	Hammer et al.[59]	https://past.en.lo4d.com/windows
Python	Van Rossum and Drake et al.[68]	https://www.python.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for original images and data should be directed to and will be fulfilled by the lead contact, Silvia Muñoz-Descalzo (silvia.munoz@ulpgc.es). Enquiries about models and code should be directed to Sabine C. Fischer (sabine.fischer@uni-wuerzburg.de).

Materials availability

This study did not generate unique reagents.

Data and code availability

- All processed image data generated in this study are publicly available as of the date of publication. These have been deposited at Zenodo. DOIs are listed in the key resources table. Microscopy data reported in this study will be shared by the lead contact upon request.
- This paper reanalyses existing, publicly available data. DOIs are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

SWISS mice were bred in house at the Universidad de Las Palmas de Gran Canaria (ULPGC) at the Institute for Biomedical and Healthcare Research (IUIBS) animal facility. All mice were housed with controlled room temperature (20°C-24°C) and relative humidity (55-72%), and a 12 h light-darkness cycle. All animal studies were conducted following National and European regulations (RD 1201/2005, Law 32/2007, EU Directive 2010/63/EU). All embryos used for this study were obtained from natural matings of virgin females of 7-12 weeks of age. The sex of embryos was not determined for the analyses conducted in this study. Embryos were classified, according to their total cell number, into early (32-64 cells), mid (65-90 cells) and late stage (more than 90 cells).

METHOD DETAILS

Experimental data

Data sets

In this work, we generated one new data set and reanalysed three existing ones. Two of the data sets were imaged in our lab and the other two were obtained from publications from the Hadjantonakis lab (Sloan Kettering Institute). For our analyses, we require for each cell the information whether it is a TE or an ICM cell, its centroid position and its expression levels of NANOG and GATA6. The cells need to be classified as NANOG+ or NANOG- as well as GATA6+ or GATA6-. Furthermore, we need a list of cells that are direct neighbours to the given cell. This happened in four steps: Imaging, Image analysis, population assignment and neighbourhood extraction. For most data sets, the first three steps were performed and described in previous publications (Table 1). The neighbourhood extraction for all data sets was done in this study.

Embryo collection

Wild-type Swiss embryos were generated by in-house breeding and natural mating. Detection of copulation plug confirmed successful mating; the resulting embryos were then considered embryonic day (E) 0.5. Embryos were isolated in M2 medium (Embryomax®).

Immunofluorescence

Embryos were prepared for immunofluorescence as previously described.[57] *Zona pellucida* was removed using acid Tyrode's. Embryos were fixed in 4% paraformaldehyde in PBS for 15 minutes, then rinsed in PBS containing 3 mg/ml polyvinylpyrolidone (PBS/PVP), permeabilised in PBS/PVP containing 0.25% Triton X-100 for 30 minutes and blocked in blocking buffer, which comprised PBS containing 0.1% BSA, 0.01% Tween 20 and 2% donkey serum. Antibodies were diluted in blocking buffer and embryos incubated at 4°C overnight. Primary antibodies were anti-NANOG (eBioscience, 1:200), anti-GATA6 (R&D Systems, 1:200). They were rinsed three times in blocking buffer for 15 minutes each and incubated in secondary antibody solution for 1 hour in the dark. Secondary antibodies labelled with Alexa fluorophores (Invitrogen) were diluted 1:1000 in blocking buffer. Embryos were then rinsed three times in blocking buffer, incubated briefly in increasing concentrations of Vectashield before mounting on glass slides with Vaseline bridges to prevent their crushing in small drops of concentrated Vectashield, and subsequently sealed with nail varnish. DAPI (1:1000) was used to stain the nuclei.

Imaging and automated image analysis for data set I

Embryos were imaged using a Zeiss LSM Zeiss LSM700 and a Plan-Apochromat 40x/1.3 Oil DIC (UV) VIS-IR M27 objective, with optical section thickness of 1 μm. Embryo images were processed and analysed as previously described.[18] All images in each imaging session were obtained

using the sequential scanning mode, with the same conditions of laser intensity, gain, and pinhole, and were processed in exactly the same way. The range indicator palette option (Zeiss AIM/ZEN software) was used to ensure that no oversaturated images were taken. The three-dimensional image stacks were segmented using MINS.[58] Cells were automatically assigned to ICM or TE. The features of the cell nuclei were extracted including the nuclear centroid and volume, together with the mean intensity of NANOG and GATA6 for each nucleus. The automatically assigned TE or ICM fate was manually checked using acquired images and ImageJ. Extreme errors (over-segmentation and pyknotic nuclei) in the segmentation were removed manually when correcting the classification of TE versus ICM.

Population assignment for data sets I and II

Depending on the thickness of a sample, confocal imaging results in fluorescence intensity decay along the z-axis. We quantified this decay for each embryo by fitting a linear regression model to the expression values of the trophectoderm cells. The slope of the regression curve then provided the factor to correct for the decay in the ICM cells.

In our setup, we mount the embryos, which results in a slight squeezing along the z-axis of the image. The correction for squeezing was performed as previously described.[18]

To determine the thresholds between high and low expression levels of NANOG and GATA6, we performed a mixture analysis with PAST[59] for each transcription factor and imaging session independently.

Neighbourhood extraction for all data sets I-VI

For the neighbourhood analyses we needed to determine the direct neighbours of each cell based on the centroid positions. To this end, we derived a Delaunay cell graph representation for each embryo in all data sets as described previously[18,31] and extracted a list of direct neighbours for each cell. As we previously shown, this is a good approximation to obtain direct neighbours.[18]

Modelling and simulations

To characterise the salt-and-pepper pattern, we generated three-dimensional artificial patterns and compared their properties to those of the experimental data. Our models range from simple rule-based models to more sophisticated approaches that include nearest neighbour or distance-based neighbour signalling. All data analysis was done in Mathematica 13.0. The model simulations were done in Python 3.7 using Jupyter notebooks as an interactive development environment.

Rule-based models

We used the nuclei centroids and cell graphs from the experimental data of all data sets and all stages and assigned positive or negative fate to each cell. For each ICM, we generated three-dimensional patterns based on the following rules:

- *Checkerboard:* One cell is assigned a positive fate. All direct neighbours of that cell are assigned a negative fate. The next cell in the list of centroids that has not been assigned a fate yet obtains a positive fate. This procedure is repeated until all cells have a positive or negative fate. In the resulting pattern, positive cells have only negative neighbours.
- *Alternating:* Starting with one centroid, a list of centroids is generated such that the next one is the closest centroid in the ICM to the previous one with respect to Euclidean distance. The cells in this list are assigned alternating positive and negative fate. This results in a pattern in which cell fate is alternating between nearest neighbour cells, i.e. if a cell is positive the cell that is closest with respect to Euclidean distance is negative and so on.
- *Local cluster with probability p*: As for the alternating pattern, we start with one centroid and generate a list of centroids in which the next one is the closest centroid in the ICM to the previous one with respect to Euclidean distance. The first cell in the list is randomly assigned a positive or negative state. The next entry in the list, i.e. the closest cell based on Euclidean distance, is assigned the same state as the previous cell with a given probability p. We analysed modelling patterns for $p = 0, 0.1, 0.2, \ldots, 0.9$ and compared their Moran's indices to those of the experimental data (Figure S10). We found the patterns for $p = 0.5$ and $p = 0.6$ to both fit some of the experimental conditions. Therefore, we further investigated the pattern for $p = 0.55$ and found it to fit the experimental data for early and mid embryos sufficiently well based on hypothesis testing (see below)
- *Random:* We extract the number of positive and negative cells from the experimental data for the given ICM. Then we randomly assign that number of positive and negative fate to the cells. This procedure is conducted both for NANOG and for GATA6, resulting in two types of patterns.

For the checkerboard rule, we did not expect the choice of the first cell to create differences in the neighbourhood properties of the pattern. Therefore, our implementation was deterministic, and we generated only one replicate per embryo. For the alternating, the local cluster and the random rule, we generated five replicates for each ICM.

Signalling models

First, we implemented a model for neighbour signalling on a fixed ICM. We used the three-dimensional embryo geometries including cell number and cell centroids from the experimental data of all data sets I-VI as the underlying ICM geometries and employed our model for transcriptional regulation based on intra- and intercellular signalling of two transcription factors u and v[25] (and below). This resulted in

three-dimensional cell fate patterns of u-v+ and u+v-, i.e. positive and negative cells (Figures 4 and 5). We performed 100 replicates per embryo geometry for each signalling model: nearest neighbour and distance-based. For the distance-based neighbour signalling, we further varied the dispersion parameter q. We chose a different value for each simulation randomly from a uniform distribution on (0,1).

As a second approach, we conducted simulations on a growing ICM (Figure 6). We started with a single cell of radius 0.75 and let the ICM evolve while the interactions for the gene regulatory network took place. Hence, each simulation time step comprised five distinct calculations:

1. Cell growth
2. Cell division
3. Adhesion and repulsion
4. Computation of the cell graph
5. Transcriptional regulation

In this case, we could not use the embryo geometries from the experimental data. Instead, we determined the average ICM cell counts of early, mid and late blastocysts as 22, 25 and 42 cells, and chose the final times for our simulations accordingly.

For the nearest neighbour signalling, we performed 200 simulations for each stage. For the distance-based neighbour signalling, we performed 1200 simulations for each stage and chose for each simulation a different value for the dispersion parameter q randomly from a uniform distribution on $(0, 1)$.

Detailed description of the modelling approach:

Cell growth

The best fitting mathematical description for cell growth is still under debate.[60] Potential mechanisms are linear[61] as well as exponential cell growth.[62] In addition, cell size control, limiting the maximum cell size might be involved.[63] More recently, it has been shown that the growth rates of cells vary between different stages of the cell cycle.[64] Due to the large number of unknowns and since the focus of our model is not on the details of cell growth, we used a simplified approach to model the growth of individual cells. We considered that cells grow by absorbing nutrients in the surrounding medium. The larger a cell becomes, the more nutrients it will be able to absorb through its surfaces, which suggests an exponential growth. However, nutrients are used to sustain the cell's own metabolism which creates a limitation to cell size. A prominent model that captures such a behaviour is the logistic growth model. We applied it to the cell radius r, i.e.

$$\frac{dr}{dt} = \lambda r(r^* - r),$$

where λ denotes a constant growth rate. The solution to this equation for initial condition $r(t_0) = r_0$ is given by

$$r(t) = \frac{r^*}{1+\left(\frac{r^* - r_0}{r_0}\right)e^{-\lambda r^*(t-t_0)}}. \qquad \text{(Equation 1)}$$

Cell division

Cell division is usually described as a function of either the elapsed time[65] or the size of a cell.[32] In the latter case, cell division has been successfully implemented for *in silico* generation of mouse ICM organoids.[32] The division itself has been described as a stochastic process of a cell's radius. Likewise, we used a stochastic approach towards cell division based on the cell's radius. We used the cumulative distribution function (CDF) $F(r)$ of a truncated normal distribution to define the probability that a cell will have divided up to some radius r. The CDF is defined via

$$F(r) = \frac{\mathrm{erf}(\rho) - \mathrm{erf}(\rho_{\min})}{\mathrm{erf}(\rho_{\max}) - \mathrm{erf}(\rho_{\min})},$$

$$\rho = \frac{r - \mu_{\mathrm{div}}}{\sqrt{2}\sigma_{\mathrm{div}}}, \rho_{\min} = \frac{r_{\min} - \mu_{\mathrm{div}}}{\sqrt{2}\sigma_{\mathrm{div}}}, \rho_{\max} = \frac{r_{\max} - \mu_{\mathrm{div}}}{\sqrt{2}\sigma_{\mathrm{div}}},$$

$$r \in [r_{\min}, r_{\max}], \qquad \text{(Equation 2)}$$

where μ_{div} and σ_{div} are the mean and standard deviation of this distribution. We chose $\mu_{\mathrm{div}} = \frac{1}{2}(r_{\min} + r_{\max})$ to preserve the symmetry of the distribution despite truncation. The function erf is the error function defined as

$$\mathrm{erf}(x) = \frac{2}{\pi}\int_0^x e^{-x^2}dx.$$

Outside of $[r_{min}, r_{max}]$, we define

$$F(r) = 0, r < r_{min},$$

$$F(r) = 1, r > r_{max}.$$

Hence, no cell divides up to a radius r_{min}, whereas cells with a radius greater than r_{max} always divide. In a time discrete process, the division probability cannot be directly calculated from Equation 2. Instead, previous attempts to divide have to be included as well. The exact division probability of a cell i with radius $r_i^{(N)}$ at time step N is then described by

$$p_i^{(N)} = \frac{F\left(r_i^{(N)}\right) - F\left(r_i^{(N-1)}\right)}{1 - F\left(r_i^{(N-1)}\right)}. \quad \text{(Equation 3)}$$

Following cell division, we use mass/volume conservation to calculate the daughter cell's radii r_1 and r_2 as

$$\frac{4}{3}\pi r^3 = \frac{4}{3}\pi r_1^3 + \frac{4}{3}\pi r_2^3.$$

In our model, we assume symmetric cell division, i.e. $r_1 = r_2 =: r_0$. This yields

$$r_0 = \frac{r}{2^{\frac{1}{3}}}. \quad \text{(Equation 4)}$$

The axis of division is randomly oriented.

Cell-cell adhesion and repulsion

The growth of a cell inevitably causes the displacement of itself and any adjacent cells. The corresponding equations of motion for n cells were obtained by an overdamped approximation, i.e. inertia was neglected such that the velocities of the cells are proportional to the forces acting on them, i.e.

$$\frac{d\mathbf{x}_i}{dt} = F_0 \sum_{\substack{j=1 \\ j \neq i}}^{n} F_{i,j} \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|}, \text{for } i = 1, \ldots, n. \quad \text{(Equation 5)}$$

Variable $\mathbf{x}_i$ describes the centroid of cell i. Parameter F_0 is a constant scaling factor, whereas $F_{i,j}$ describes the magnitude of the force of cell j acting on cell i. The forces are derived from the Morse potential, which has already been successfully applied in similar biological contexts.[66,67] They can be written in terms of cell radii r_i, r_j and positions $\mathbf{x}_i$, $\mathbf{x}_j$ such that

$$F_{i,j} = \begin{cases} 2\alpha\left(e^{-\alpha\left(|\mathbf{x}_j - \mathbf{x}_i| - \sigma\left(r_j + r_i\right)\right)} - e^{-2\alpha\left(|\mathbf{x}_j - \mathbf{x}_i| - \sigma\left(r_j + r_i\right)\right)}\right) & \text{for } |\mathbf{x}_j - \mathbf{x}_i| \leq r_j + r_i, \\ 0 \quad \text{for } |\mathbf{x}_j - \mathbf{x}_i| > r_j + r_i. \end{cases}$$

The parameter α describes the stiffness of the cells, whereas $\sigma \in (0, 1]$ defines the optimal distance between two cells in contact as a fraction of the sum of their radii. The effect of the forces depends on the relative position of two cells. If the cells are too close, they will repel (Figure S16A (i)), whereas they will adhere to each other if they are too far from each other but still in contact (Figure S16A (ii)). In between, an optimal state will be found (Figure S16A (iii)). If the distance between the cells exceeds the sum of their radii, there can be no physical interaction (Figure S16A (iv)).

Transcriptional regulation

We employ our previously developed model for transcriptional regulation.[25] It follows the temporal evolution of two transcription factors u and v based on a model derived from statistical thermodynamics. The corresponding gene regulatory network (GRN) (Figure S16B) involves mutual inhibition of u and v as well as their auto-activation. An external signal s is used to inhibit u while also activating v. For an ICM with n cells, we end up with a system of ordinary differential equations (ODEs)

$$\frac{du_i}{dt} = r_u \frac{\eta_u u_i}{1 + \eta_v v_i (1 + \eta_s \eta_{vs} s_i) + \eta_u u_i + \eta_s s_i} - \gamma_u u_i,$$

$$\frac{dv_i}{dt} = r_v \frac{\eta_v v_i(1+\eta_s \eta_{vs} s_i)}{1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i} - \gamma_u v_i,$$

$$\text{for } i = 1, \ldots, n, \qquad \text{(Equation 6)}$$

with the energy coefficients $\eta_x = e^{-\Delta\epsilon_x}$, the transcription rates r_x and the decay rates γ_x.

We consider two types of external signal s. For the nearest neighbour signalling that is activated by u, we consider

$$s_i = \frac{1}{|N_G(i)|} \sum_{j \in N_G(i)} u_j. \qquad \text{(Equation 7)}$$

Here, we used the notation $N_G(i)$ from graph theory to denote the neighbours of vertex i in the graph G.

For the distance-based neighbour signal, s is given as the result of the expressions of u in any other cell according to

$$s_i = \frac{1}{\max_k \sum_{j \neq k} q^{d_{kj}-1}} \sum_{j \neq i} u_j q^{d_{ij}-1}, q \in [0, 1]. \qquad \text{(Equation 8)}$$

Here, d_{ij} denotes the distance of cells i and j in the cell graph. By this definition of s_i, the parameter q represents the strength of signal dispersion throughout the ICM. A value of $q = 0$ means that only direct neighbors influence a cell's fate. Increasing q increases the dispersion range of the signal.

If transcriptional regulation is coupled with cell division, the values of u_i and v_i are evenly distributed between the daughter cells upon division.

Implementation

The implementation of the computational model is realized by a time discrete process. Beginning with a single cell, in each time step, the above-mentioned steps are carried out sequentially. First the cell radii are updated via Equation 1. Afterwards, the cell division probability is calculated according to Equation 3. For each cell, we draw a random number from a uniform distribution between 0 and 1. If this number is less than the division probability, the cell is chosen for cell division. Thus, two cells are generated with equal radii according to Equation 4. The axis of division is chosen randomly. The distance of the two cells after cell division is fixed as a parameter h. The amounts of u and v are equally distributed between the two daughter cells. Afterwards, the cell displacement is implemented by solving a single step of Equation 5 numerically using the explicit Euler scheme. Finally, the transcriptional regulation is implemented by a second step of the explicit Euler scheme applied to Equation 6 with Equation 7 or Equation 8 as also described in.[25] The following tables further specify the variables and parameters used in the simulation (Tables 2 and 3). The evolution times were fitted to match average ICM cell counts of early (22 cells), mid (25 cells) and late (42 cells) blastocysts.

QUANTIFICATION AND STATISTICAL ANALYSES

Moran's index is a measure to quantify spatial autocorrelations. Its original form was first introduced in[27] to analyse the distribution of soil fertility over a field. For a variable $x \in \mathbb{R}^n$, we calculate Moran's index as

$$I = \frac{n}{\sum_{ij} A_{ij}} \frac{(x - \bar{x})^T A (x - \bar{x})}{(x - \bar{x})^T (x - \bar{x})}$$

where x includes the fates of the n cells, $\bar{x}$ denotes the mean of x and A is the adjacency matrix of our Delaunay cell graph. Considering that we only incorporate two cell fates in our study, namely positive and negative, x is given by a vector with entries $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$. The value of I is independent of which value is assigned to which fate.

To get an intuition of Moran's index as a measure of spatial patterns, we consider two extreme cases in a synthetic two-dimensional example on an 8 x 8 grid (Figure S16C). In a checkerboard pattern, equal cells are never adjacent to each other as long as the number of neighbours does not exceed four. Thus, there is no auto-correlation between the cells leading to the minimal value $I = -1$. In the case of spatial separation, cells are mostly neighbours to equal cell types resulting in values of I close to 1.

In the **Box-Whisker-Chart**, the boxes denote the range from 25% to 75% quantile. The white line indicates the median. Values that are larger than the 75% quantile plus 1.5 times the interquartile range are near outliers, marked in black. Values that are larger than the 75% quantile plus three times the interquartile range are far outliers, marked in grey.

ICMs that contain only one cell type were excluded from the analysis. We excluded 51 early, 2 mid and 2 late embryos for NANOG, and 64 early, 19 mid and 6 late embryos for GATA6.

For the simulation results of the distance-based neighbour signalling model, we displayed Moran's index as a function of q. To this end, results were binned according to their q values in bins of size 0.1.

To determine the best fitting model, we compared the model and simulations results with the experimental results both visually and by employing **hypothesis testing**. We used Welch's t-test with Bonferroni correction and a significance level of 0.05, unless otherwise stated (Figure 3E).