Manuscript submitted to **Biophysical** *Journal*
Article

Cell Fate Clusters in ICM Organoids Arise from Cell Fate Heredity & Division – a Modelling Approach

T. Liebisch[1,*], A. Drusko[1], B. Mathew[2], E. H. K. Stelzer[2], S. C. Fischer[3], and F. Matthäus[1]

ABSTRACT

During the mammalian preimplantation phase, cells undergo two subsequent cell fate decisions. During the first cell fate decision, cells become either part of an outer trophectoderm or part of the inner cell mass. Subsequently, the inner cell mass segregates into an embryonic or an extraembryonic lineage giving rise to the epiblast and the primitive endoderm, respectively. Inner cell mass organoids represent an experimental model system for preimplantation development, mimicking the second cell fate decision taking place in *in vivo* mouse embryos. In a previous study, the spatial pattern of the different cell lineage types was investigated. The study revealed that cells of the same fate tended to cluster stronger than expected for purely random cell fate decisions. In order to investigate the emergence of the cell lineage type clustering behaviour, we developed an agent-based model. Hereby, cells are mechanically interacting with direct neighbours, and exert adhesion and repulsion forces. The model was applied to compare two current assumptions of how inner cell mass neighbourhood structures are generated. We tested how different assumptions regarding cell fate switches affect the observed neighbourhood relationships. The model supports the hypothesis that initial cell fate acquisition is a stochastically driven process, taking place in the early development of inner cell mass organoids. The model further shows that the observed neighbourhood structures can emerge due to cell fate heredity during cell division and allows the inference for a time point for the cell fate decision.

STATEMENT OF SIGNIFICANCE

Cell fate decisions in early embryogenesis have been considered random events, which also cause a random distribution of cells of different cell fates. Using an agent-based mathematical model, fitted to data derived from ICM organoids, we show that a random distribution occurs only for a short time interval, as cell fate heredity and cell division quickly lead to spatial cell fate clustering. Our results speak against neighbourhood interactions determining an individual cell fate. Instead our approach indicates four consecutive phases of early development: 1) co-expression of cell fate markers, 2) cell fate decision, 3) cell division and local cell fate clustering, and 4) phase separation, whereby only the phases 1-3 occur in ICM organoids during the first 24h of growth.

INTRODUCTION

The first steps during mammalian embryo development are ovulation and fertilisation, followed by the preimplantation phase. At this point, the blastocyst is formed, which later implants into the uterus (1). Postimplantation development rapidly proceeds and involves multiple cell differentiation and morphological changes (1, 2). The first steps within the complex development processes in mammalian systems involve the cell fate decisions during the preimplantation phase. During development, the preimplantation phase is key to improved pregnancy for mammals (3). Despite the importance of preimplantation, processes taking place during this phase are not fully understood.

The mouse is a common model organism to study the preimplantation phase. The 8 to 16 cells morula is formed until the embryonic day 2.5 (E2.5) after fertilisation. At this stage, the first cell fate decision is taking place. Cells on the surface of the morula become trophectoderm (TE) while cells inside the morula are forming the inner cell mass (ICM) (4, 5). During E2.5 to E4.5 the second cell fate decision process takes place: ICM segregates into epiblast (Epi) and primitive endoderm (PrE) (6–8).

NANOG and GATA6 are described as the first markers for Epi and PrE segregation, respectively. Expression levels of NANOG and GATA6 undergo progressive changes during the morula stage and the early blastocyst (9, 10). In early blastocysts (E3.0), all ICM cells seem to co-express NANOG and GATA6 (7, 11, 12). Subsequently, NANOG and GATA6 are gradually up- or down-regulated during the 32-cell stage. Thereby, both transcription factors repress each other locally (10, 13–17), leading to a mutually exclusive transcription factor expression in late blastocysts (64 cells) (18). Once a cell-fate is determined it is only possible to switch the fate by an external modulation of the included signalling pathways (16, 17, 19).

Several recent models suggest different hypotheses for cell fate decisions from early mouse blastocyst to mid mouse blastocyst (20–23). Those models suggest either a bistable (20) or a tristable (21, 22) regulatory network in the ICM. Tosenberger *et al.* (2017) implemented a multi-scale model taking account of cell division and cell movement on the one hand, and a gene regulatory network on the other hand. They aim to explain initial cell fate assignment by intracellular regulatory networks formed by the transcription factors NANOG and GATA6 and the FGF/ERK pathway.

While at early phases of preimplantation the spatial distribution of NANOG and GATA6 positive cells is commonly described as a random salt-and-pepper pattern, GATA6 positive cells are sorted to the rim of the ICM in the late blastocyst while NANOG positive cells are forming an inner core (1, 7, 10, 11, 14). Over time, two major hypothesis were developed to explain the development of the spatial ICM pattern, with NANOG positive cells positioned at the centre of the ICM and GATA6 positive cells forming the outer layer. (1) The more modern *cell sorting model* states that the ICM is composed of cells that already are progenitor cells of Epi and PrE. Those progenitor cells give rise to their respective cell type but are not allowed to change their expression type. In a sorting phase, the different cell types segregate into the described layers. This sorting process is described to be tightly connected to cell divisions and the geometry of the blastocyst (7, 24–26); (2) the older *bipotent model* states that cells in the ICM are homogeneous as well as bipotent and thus might segregate into the different cell types. Segregation is claimed to be initialised by a hypothetical inductive signal from outside the ICM (6, 27, 28).

A recent study introduced a novel 3D stem cell system named *ICM organoids* (in the following the term organoid refers to the biological system while the term spheroid is used in context of modelled data), which is based on inducible mouse embryonic stem cells (29). ICM organoids mimic the segregation into Epi and PrE without forming a TE and show key events and timing of cell fate specification and cell-cycle progression in the ICM of mouse blastocysts. Thus, ICM organoids provide a powerful tool to develop and test biological preimplantation hypotheses *in vitro*. At 24 h of growth, most of ICM organoid cells express either NANOG or GATA6. The spatial segregation into an inner core of NANOG expressing cells and an outer layer of GATA6 expressing cells is visible after 48 h of growth. However, in contrast to mid mouse blastocysts, which consist of approximately 64 cells, ICM organoids comprise over 400 cells after 24 h of growth and more than 1000 cells after 48 h of growth.

In order to quantify the patterns of neighbourhood distributions, a neighbourhood analysis of NANOG, GATA6, double positive and double negative cells was conducted in ICM organoids (29). Interestingly, the study revealed a local clustering of cells sharing the same expression type, inconsistent with the assumed salt-and-pepper pattern of cell fates. Different static models were tested to determine whether the 3D-neighbourhood pattern of ICM organoids in the state of mutual exclusive expression could be explained by simple stochastic rules. Three of the presented simulations relied on a random pattern, while the fourth simulation was based on a local cell fate clustering. However, none of the tested stochastic rules could reproduce the observed neighbourhood pattern (29).

In order to investigate the emergence of the observed 3D-neighbourhood structure in 24 h old ICM organoids, we use a 3D agent-based model. Agent-based models provide a technique to represent a wet-lab experiment under idealised conditions (30). The model is given as a set of differential equations, describing mechanical cell-cell interactions, such as adhesion and repulsion forces, stochastic cell fate decision (omitting a detailed description of the signalling pathway dynamics), cell growth, and cell division involving cell fate heredity. Comparable agent-based models are commonly used to study cancer growth, cell proliferation or the contribution of single cells towards collective cell migrations (23, 31–35).

We use the model to investigate the hypothesis that the observed cell fate clustering arises from cell divisions, whereby cell fates are (partially) passed on to both daughter cells. Simulations were conducted under two hypotheses, during cell division, the cell fate is passed on to both daughter cells, switches between different fates are not allowed (H1); during cell division, double-positive cells may give rise to double positive or single positive cells. Thus, allowing cell fate switches of double positive cells to single positive (H2).

Our results indicate that the observed cell fate clustering can indeed be explained as a randomly distributed cell fate decision with subsequent divisions and cell fate heredity. Hereby, the simulations performed under H1 reflect the neighbourhood statistics better than the simulations performed under H2. Furthermore, based on the neighbourhood statistics, a time point for the cell fate decision (prior to the 24 h growth stage) can be inferred.

MATERIALS AND METHODS

Experimental Data

Our study is based on experimental results from Mathew *et al.* (2019). Here, expression levels for NANOG and GATA6, as well as the nuclei positions for all cells in 24 h and 48 h old ICM organoids were determined by fluorescence intensity measurements. In total, four different expression types were established: NANOG positive and GATA6 positive (double positive | N+G+), NANOG negative and GATA6 negative (double negative | N-G-), NANOG positive and GATA6 negative (N+G-), NANOG negative and GATA6 positive (N-G+).

In the following, the stage of 24 h old and 48 h old ICM organoids will be referred to as t_1 and t_2, respectively. During the simulations, cell fates and positions were recorded when the cell count in the simulation coincided with the average ICM organoid cell count in the experiment at t_1 (441 cells at 24 h) and t_2 (1041 cells at 48 h).

Model Implementation

An individual cell-based model that defines a small set of cellular features, implemented by Stichel *et al.* (2017) , is extended to explain the rise of local cell fate clustering in ICM organoids. The model describes the displacement of cells in response to external forces exerted by surrounding cells:

$$\mathbf{F}_{i,k} = F_0 \cdot F(r_i, r_k, ||\mathbf{x}_i - \mathbf{x}_k||) \cdot \frac{\mathbf{x}_i - \mathbf{x}_k}{||\mathbf{x}_i - \mathbf{x}_k||}, \tag{1}$$

with F_0 a positive constant, representing the strength of the mechanical interaction, r_i the radius of the ith cell, $\mathbf{x}_i = (x_i, y_i, z_i)$ the position of the ith cell and

$$F(r_i, r_k, d) = \begin{cases} 2 \cdot (e^{-2a(d-(r_i+r_k))} - e^{-a(d-(r_i+r_k))}) & d < \sigma r_i \\ 0 & d \geq \sigma r_i \end{cases}. \tag{2}$$

As given by the Morse potential (Eq. 2), the force between two cells is positive (repulsive) if the distance between the cell centres is below the sum of their radii, and negative (attractive) for $(r_i + r_k) < d < \sigma r_i$. Repulsion accounts for the limited compressibility of cells, while attraction accounts for cell-cell adhesion. The attractive part of the potential is cut off for cells at a distance above σr_i. If the distance of two cells equals the sum of their radii they are in perfect distance and thus neither emit attraction nor repulsion onto each other. Displacement of cells in this model is only determined by forces exerted on them:

$$\frac{d\mathbf{x}_i}{dt} = \sum_{k,k\neq i} \mathbf{F}_{i,k}. \tag{3}$$

Each cell in this model is described by three features. A position $\mathbf{x}$, a radius r and an expression type ϵ. Other model parameters are assumed to be the same for all cells (e.g. elasticity, adhesion strength). The radius (size) of the cells is growing over time with

$$\frac{dr_i}{dt} = k \cdot (r^* - r_i), \tag{4}$$

with k a (positive) growth constant and the maximum cell size r^*. Cell division is determined by a stochastic process which depends on the cell radius but not on the cell type. During cell division, the cell volume is preserved. The mother cell keeps its position ($\mathbf{x}_m$) and reduces its radius (r_m) with:

$$r_{m,new} = r_m * \sqrt[3]{\frac{1}{2}} \tag{5}$$

The daughter cell ($\mathbf{x}_d$) is generated close to $\mathbf{x}_m$, with $\mathbf{x}_d = \mathbf{x}_m + \mathbf{E}$ with $\mathbf{E}$ a random 3D vector containing small values. The daughter cell is assigned to the same size as the mother cell ($r_d = r_{m,new}$). Directly after cell division, both cells are growing with Eq. 4 and changing their positions with Eq. 1.

Since it was shown that the initial cell fate decision can be described as a stochastic process (36–38), the initial expression type ϵ of the simulated cells is assigned randomly from the four expression types

$$\epsilon \in \{\mathrm{N\text{-}G\text{-}}, \mathrm{N_+G_+}, \mathrm{N_+G\text{-}}, \mathrm{N\text{-}G_+}\}. \tag{6}$$

Both assumptions, H1 as well as H2, can be tested using our simulations. We assign the initial cell fate randomly and fit the cell fate probabilities to match the cell type proportion data for t_1. For the fit, the dynamics of the heredity of cell fates during cell division are formalised using a set of ordinary differential equations. Using experimental data for 24 h old ICM organoids as known result, the unknown initial conditions (expression type proportions for initial cell fate assignment) are identified using the Euler Backward method. To test the assumption H1, cells are only allowed to pass on their own expression type to daughter cells during cell division (i.e. Eq. 7).

$$\begin{pmatrix} d\mathrm{N_+G_+}/dt \\ d\mathrm{N\text{-}G\text{-}}/dt \\ d\mathrm{N_+G\text{-}}/dt \\ d\mathrm{N\text{-}G_+}/dt \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_4 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{N_+G_+} \\ \mathrm{N\text{-}G\text{-}} \\ \mathrm{N_+G\text{-}} \\ \mathrm{N\text{-}G_+} \end{pmatrix}, \tag{7}$$

with N-G-, $\mathrm{N_+G_+}$, $\mathrm{N_+G\text{-}}$, and $\mathrm{N\text{-}G_+}$ representing the cell count of cells with a given cell fate and α representing the probability of a cell division event. We assume, that cells of the different cell fates show the same cell division rates (i.e. $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4. = 1$). For the assumption H2, cells of the expression type $\mathrm{N_+G_+}$ are assumed to be bipotent, and thus are allowed to change their expression type during cell division. However, for this model the observed proportion data for t_2 could only be matched when N+G+ cells produced N-G+ cells during cell division (i.e. Eq. 8). Eq. 7 and 8 do not account for the spatial component, but assume a well mixed system.

$$\begin{pmatrix} d\mathrm{N_+G_+}/dt \\ d\mathrm{N\text{-}G\text{-}}/dt \\ d\mathrm{N_+G\text{-}}/dt \\ d\mathrm{N\text{-}G_+}/dt \end{pmatrix} = \begin{pmatrix} \alpha_1 - \beta & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ \beta & 0 & 0 & \alpha_4 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{N_+G_+} \\ \mathrm{N\text{-}G\text{-}} \\ \mathrm{N_+G\text{-}} \\ \mathrm{N\text{-}G_+} \end{pmatrix} \tag{8}$$

Given the experimentally determined cell fate proportions for 24 h and 48 h old ICM organoids, other cell fate switches have not been feasible. For both assumptions, the initial cell fate proportions have been determined and assigned to take place at different time points (t_0). In particular, we couple t_0 to the cell count and initiate the cell fate when the simulated ICM spheroids reached a cell count of 75, 100, 150, 200, 250, 300 or 400 cells. Each simulation has been repeated 100 times.

Simulations were performed with cell fate proportions for an ICM spheroid comprising a particular amount of cells (t_0) given as input. Therefore, we used Eq. 7 or 8 to determine cell proportions at each individual cell count for t_0. Subsequently, a stochastic simulation was performed and results for cell positions and cell fates at t_1, t_2 have been saved. The proportions measured at t_1 should represent those measured in 24 h old ICM organoids. Using the position and cell fate data of cells, their neighbourhood structures at t_1 and t_2 have been determined using a 3D Delaunay triangulation and a distance-based cutoff (σr_i).

Neighbourhood analysis and statistics

Cell neighbours were determined for both, simulation and experimental data, using Delaunay triangulation. For the neighbourhood statistics, we derived the set of all neighbours of all cells of a given fate j, and computed the proportion of cell types based on this neighbourhood set. Neighbourhood proportions were collected for every executed simulation. For statistical comparison between experimental and simulated data according to single neighbourhood structures, we used the Wilcoxon-Mann-Whitney test with Bonferroni correction for multiple testing. In order to compare the overall fit of the simulated pattern to experimental data, we used the effect size as the relative deviation (ψ) as given by Mathew *et al.* (2019).

$$= \frac{|(\bar{s} - \bar{m})|}{\bar{m}}, \tag{9}$$

with $\bar{s}$ representing the mean of simulated data and $\bar{m}$ the mean of the experimental data.

Spatial analysis of expression type distribution

During visual inspection of biological data, we noticed that double negative cells often clustered at the rim of the ICM organoids. In order to visualise this effect, the following procedure was applied to experimental data for 24 h old and 48 h old ICM

organoids. The spatial heterogeneity is assumed to be the same in small and in large ICM organoids, thus we normalised the size of all ICM organoids to equalise them in space. For normalisation, the median absolute divergence is used instead of the standard deviation because it is more robust with respect to outliers. The centre of mass for double negative cells was determined and the entire ICM organoid was rotated such that the centre of mass was located on the x-axis (i.e. $x > 0, y = 0, z = 0$). The rotated cell positions were combined in one data set. Each double negative cell was assigned the value 1 and the value 0 was assigned to all other cells. Interpolation on the combined data generated a continuous clustering pattern from experimental data. In the interpolated dataset, the label 1 indicated the presence of a particular expression type, while 0 indicated its absence. The procedure was repeated for double positive, NANOG positive and GATA6 positive cells to show that the visualised heterogeneity is not an artefact resulting of this procedure.

In order to analyse the cell fate proportions in dependence on their relative distance to the ICM organoid centre, we measured the distance of the cells of each ICM organoid to their respective centre of mass. Subsequently we divided these distances into 10 intervals. Eventually the mean proportions and standard deviations for the 10 intervals were determined for all 24 h old ICM organoids and 48 h old ICM organoids. Unless otherwise stated, the model and data analysis methods were implemented in MATLAB R2019a.

RESULTS

The purpose of the simulations was to explain the complex 3D neighbourhood pattern by simple mechanisms, based on cell mechanics, a stochastic expression type assignment and cell fate heredity during cell division. Furthermore, the simulations aimed to predict the ICM organoid cell count when the cell fate decision occurs. For this reason, we used a mechanistic mathematical model for 3D spheroid growth in order to investigate the distribution pattern of cell fates in 24 h old ICM organoids (see Methods). During the simulation, the cells were assigned to the different expression types (i.e. N-G-, N+G+, N+G- and N-G+) randomly at t_0, whereby the cell type proportions were set to fit the experimental data for 24h old ICM organoids. The time point for the initial expression type assignment (t_0) was coupled to cell count of the spheroid (75, 100, 150, 200, 250, 300 or 400 cells). One 24 h old ICM organoid, one 48 h old ICM organoid, as well as simulated ICM spheroids for t_1 and t_2 are shown in Fig. 1 A.

We tested two hypotheses addressing different cell fate heredity strategies (see Introduction). In assumption H1, each cell passes on its cell fate to both daughter cells. For assumption H2, double-positive cells are allowed to give rise to double-positive or GATA6 positive cells (see Fig. A3). The production of GATA6 cells from double-positive cells is the most reasonable assumption for any cell fate switch because a) a switch from NANOG to GATA6 or vice versa is biologically not feasible (16, 17), b) the amount of double-positive cells remains constant between the time points t_1 and t_2, and c) the proportion of GATA6 cells increases strongly from t_1 to t_2.

Starting from the initial cell fate assignment at t_0, cell growth, displacement according to adhesion and repulsion forces, and cell divisions were simulated according to Eqs. (1)-(6). Cell fates and positions were recorded at t_1 and t_2.

The expression type compositions of simulated ICM spheroids were determined and statistically compared with experimental data using the Wilcoxon-Mann-Whitney test with Bonferroni correction. For the assumption H1 (Fig. 1 B) and also for the assumption H2 (Fig. 1 C), the simulated proportions agreed very well with the experimental data at t_1. At t_2 the simulated proportion data showed for the assumption H1 significant differences to experimental data for all expression types at all cell fate initialisation time points t_0 ($p < 0.05$). For the assumption H2 (Fig. 1 C), significant differences between predicted and experimental data at t_2 were determined for cells with the expression types N-G-, N+G- and N-G+ ($p<0.05$). No statistically significant differences were found between simulated and experimental data for cells with the expression type N+G+ at t_2.

The neighbourhood distribution for the assumption H1 (see Fig. 1 D) largely agreed with experimental data at t_1, the only exception being the neighbourhood statistics involving N-G- cells. Overall, the neighbourhood distributions fitted the experimental data best, if the cell fate assignment occurred at $t_0 = 150$ (lowest ψ-value). Concerning the neighbourhood structures measured at t_2, the prediction power of the model decreased strongly (see Fig. A3 A). Only the neighbourhood structures of N+G- cells showed reasonably good agreement for t_0 lower than 350 and larger than 100 cells.

According to the overall comparison of neighbourhood structures, the best agreement between experimental data and simulation required $t_0 = 150$ cells. The statistical comparison of the explicit neighbourhood statistics has been performed using the Wilcoxon-Mann-Whitney test. Determined p-values for simulations with $t_0 = 150$ cells are shown in the Supplementary Material for t_1 (see Tab. A1, A3).

The neighbourhood statistics for the assumption H2 agreed reasonably well with the experimental data (see Fig. 1 E). The neighbourhood pattern of cells with the expression types N-G- and N+G- at t_1 were similar to the neighbourhood structures obtained under assumption H1, including the misfit involving N-G- cells. In addition, simulated N+G+ cells were significantly less often neighbours of other N+G+ cells than the experimental data suggests. The latter speaks against a cell fate change of N+G+ cells into GATA6 cells.

Evaluated patterns for t_2 showed a better agreement with the experimental data, performed under the assumption of the H2 model. In particular, the predicted neighbourhood statistics to double positive cells were statistically less significant between the *in vitro* measured patterns and the simulated cell neighbourhood statistics under the assumption H1. The simulated neighbourhood structures are shown in the Supplementary Material (see. Fig A3). The determined p-values for simulations with $t_0 = 150$ cells are shown in the Supplementary Material for t_2 (see Tab. A2, A4).

Spatial analysis of expression type distribution

Simulation results of both model hypotheses fitted well to the experimental data on the spheroid expression type composition, and could also explain the neighbourhood statistics to a large degree. The only disagreement between simulation and experiment concerned the neighbourhood statistics of N-G- cells. In order to investigate the reasons for this disagreement, we conducted an additional spatial analysis of the expression type distribution. Fig. 2 shows 3D views of the average ICM organoid composition for both time points t_1 and t_2, marking the spatial density of a given cell fate j (see Methods). Cells with the expression type N+G+ were spread evenly over the whole ICM organoid at t_1 and t_2. The same distribution pattern was obtained for N+G- and N-G+ expression type cells at t_1. Their spatial distribution pattern changed over time. At t_2, N+G- cells formed a cluster in the centre of the ICM organoid while N-G+ cells formed an evenly distributed outer layer around the inner core of the ICM organoid. Concerning the N-G- cells, we find that interestingly, this expression type tended to be positioned in the outer parts of the ICM organoid at both time points, t_1 and t_2. In both cases their distribution was unevenly spread over the outermost layer of cells in the ICM organoid.

DISCUSSION

In this study, we investigated the neighbourhood structures of 24 h and 48 h old ICM organoids with mutual exclusive expression, resembling the mid mouse blastocysts (E3.75). We demonstrated that a simple model, involving mechanical interactions such as adhesion and repulsion, cell division with cell fate heredity, and a stochastically driven cell fate, can explain the complex 3D spatial arrangements of the neighbourhood statistics in 24 h old ICM organoids. The proposed model is based on a minimal number of assumptions: (1) Up to a certain time point, when the simulated ICM spheroid consists of a predetermined number of cells (i.e. t_0), all cells coexpress the transcription factors NANOG and GATA6, analogous to mouse embryonic development to E3.0 (11, 37). (2) At t_0 a cell fate is introduced to each cell with a given probability (omitting a detailed description of the signalling pathway dynamics) (37, 38). Cells either continue to co-express both transcription factors (double positive), stop the expression of one of them (NANOG or GATA6 positive), or stop to express both (double negative). The probabilities for each case are extracted from experimental data (29). (3) Once a certain cell fate has been chosen cells do not switch to another cell fate (21). (4) During cell division, the cell fate is passed on to the daughter cells according to certain rules (see Fig. A3). Two hypotheses addressing the passing on of the cell fate during cell division were tested. During cell division the cell fate is passed on to both daughter cells, switches between different fates are not allowed (H1); during cell division double-positive cells may give rise to double positive or single positive cells. Thus, allowing cell fate switches of double positive cells to single positive (H2).

Proportions

We used the setup to compare two hypotheses. For H1, cell fate switches are not allowed during cell division and for H2, cell fate switches are allowed during cell division (6, 7, 24, 25, 27, 28). Since the initial cell fate assignment at different t_0 was adjusted to fit the proportion data known from t_1, both models reproduced the proportions for t_1 very well. The only exception was that the assumption H2 could not reproduce the proportion data if $t_0 \leq 150$.

At t_2, GATA6 expressing cells represented the majority in the ICM organoid. This implies an enhanced cell division rate for GATA6 positive cells, for instance due to an interaction between the cell growth rate and mechanical forces exerted by surrounding cells. Cells on the surface of ICM organoids might grow faster than cells in the centre of ICM organoids, where the cell density is high (31, 39). GATA6 expressing cells are predominantly found at the rim of ICM organoids where the cell density is lower, which might allow a higher cell division rate (40, 41). A feedback of pressure onto cell growth is not included in our model, hence this aspect can not be captured.

Neighbourhood Structures

Complex 3D spatial arrangements of cells with particular cell fates characterise early stages of PrE and Epi segregation. ICM organoids mimic those cell arrangements after 24 h of growth and after 48 h of growth for mid and late mouse blastocysts, respectively. Mathew *et al.* (2019) used a computational rule-based static model in order to reproduce the cell fate pattern

measured of 24 h old ICM organoids. Hereby, four different hypotheses have been tested. The first three simulations were based on hypotheses derived from the salt-and-pepper pattern, while the fourth tested pattern was based on a local cell fate clustering. Patterns from all four simulations were significantly different from experimentally observed patterns (29), in particular, because the clustering of the cells sharing the same cell fate could not be reconciled with any of the tested models.

In this study, we showed that a salt-and-pepper distributed cell fate decision, in addition to cell division involving cell fate heredity, adhesion and mechanical repulsion, could explain the lineage composition and spatial distribution of GATA6 and NANOG in 24 h old ICM organoids. The main assumption here was that the 24 h time point did not represent the time point of the cell fate decision. Instead, we assumed that the cell fate decision occurred previously, and that a local clustering of cells of the same fate arises from cell divisions. Under this hypothesis we found that both, the assumption H1 as well as the assumption H2, showed large agreement with experimental data for 24 h old ICM organoids, whereby the assumption H1 reflected the neighbourhood structure of double positive cells better than the assumption H2.

A comparable model, which is also based on mechanic cell displacement and cell division, but takes also into account an intercellular network for cell fate decisions was used by Tosenberger *et al.* (2017) to predict cell fate decisions in *in vivo* mouse blastocysts. According to their findings, the spatio-temporal distribution of PrE and Epi cells follows the salt-and-pepper pattern, whereby Epi cells are preferentially surrounded by PrE cells and vice-versa (23). In general, the idea of a salt-and-pepper distributed cell fate pattern is questioned by several recent studies that are highlighting the need of neighbourhood interactions for the rise of Epi and PrE lineages (21–23). However, we hypothesise that the local spatial clustering of cell fates as obtained by Mathews *et al.* 2019 might not be indicative of specific neighbourhood interactions involved in cell fate decision. We further hypothesise that a cell fate clustering should be a transient state characteristic of blastocysts and ICM organoids in the time span between cell fate decision and spatial sorting. In particular, our results indicate that an initial cell fate decision based on a random distribution, in addition to cell divisions and cell fate heredity, can give rise to local cell clusters.

With the given restrictions on the cell fate switches, it was not possible to fit the proportion transition from 24 h to 48 h data perfectly well. Although cell fate switches are considered unlikely, we investigated if the proportion data and neighbourhood structures could be approximated better if further cell fate switches are allowed. Indeed, this relaxation allows for a better fit of the cell proportion data. However, the neighbourhood statistics are not approximated well under this new hypothesis H3 (see Supplementary Material).

Finally, cell fate clustering as a result of cell divisions implies that the cell fate decision occurred at an earlier time point. Using our model we can provide an estimate for the time point when the cell fate decision in ICM organoids should take place to yield the observed clustering. In particular, we observe the best agreement between model and data when we assume an initial cell fate assignment at a cell count of about 150 cells. Furthermore, it is possible that the cell fate decision process ranges over a longer period of time. This can be added to the model, but requires additional parameters.

Spatial analysis of expression type distribution

NANOG and GATA6 are described as the first markers for Epi and PrE cell fate segregation, with a salt-and-pepper like occurrence in early and mid mouse blastocyst. During blastocyst growth, PrE cells are sorted to the rim of the ICM, while Epi cells remain in the centre of the ICM (1, 7, 10–17). This behaviour of PrE and Epi cells is reflected in ICM organoids. While NANOG and GATA6 positive cells are evenly distributed over the 24 h old ICM organoid they re-localise in 48 h old ICM organoids, forming a centre of NANOG positive cells and a rim of GATA6 positive cells. Double positive cells are co-expressing NANOG and GATA6, as described for early mouse blastocysts (E3.0) (7, 11, 12). Thus we expected them to be distributed evenly over the whole ICM for 24 h old and 48 h old ICM organoids, which was confirmed by the conducted spatial analysis. For double negative cells, expressing neither NANOG nor GATA6, we also expected a spatially homogeneous distribution. However, the spatial analysis revealed that instead double negative cells tended to be positioned at the rim of ICM organoids at the 24 h time point. This finding was surprising, but it explains the very high proportion of double-negative cells in the neighbourhood of double-negative cells, which cannot be captured by our simulations. The simulated ICM spheroids are homogeneously composed. Thus, they do not match with the heterogenous clustering of double negative cells found in ICM organoids. We expect that the double negative cells have been GATA6 positive cells that have evolved further. However, the role and dynamics of double negative cells is not yet well understood, and should be investigated in further studies.

Conclusions

Both hypotheses (H1 and H2) are not capable of explaining the exact cell division dynamics for the transition from 24 h to 48 h old ICM organoids. Although the simulated neighbourhood structures fit parts of the experimental data reasonably well, the model is lacking a cell sorting process, which we hypothesise to result in the measured neighbourhood structures. Cells placed at the rim of cell colonies are described to grow faster, and thus show a higher cell division rate (31, 39). After the sorting of

GATA6 positive cells to the rim of the ICM organoid, these cells might divide with a higher rate, which would explain the strong increase of their proportions from 24 h to 48 h old ICM organoids.

The model hypothesises a temporal separation of four phases (see Fig. 3). At first all cells coexpress the transcription factors GATA6 and NANOG, subsequently cells are assigned to a certain cell fate. Cell division in combination with cell fate heredity then lead to a local clustering of cells sharing the same cell fate found in 24 h old ICM organoids. Eventually a cell sorting process leads to the phase separation found 48 h old ICM organoids. We expect that the phases overlap in real systems, but the good agreement between the model and data suggests that the four phases do not occur simultaneously. Although the model is lacking a cell sorting process, it captures the neighbourhood statistics of 24 h old ICM organoids, which indicates that the cell sorting process initiates later.

AUTHOR CONTRIBUTIONS

Conception and design of the study were performed by T.L. and F.M.; Experimental data were provided by B.M., S.C.F. and E.H.K.S.; T.L. implemented the code.; T.L., A.D. and F.M. analysed the data; T.L., F.M. and SC.F. interpreted the data. T.L. wrote the manuscript. All authors revised the manuscript. F.M. supervised the study.

ACKNOWLEDGMENTS

Research in the F.M. lab is supported by the Giersch foundation. F.M. and E.H.K.S. are supported by a grant from the Hessen State Ministry for Higher Education, Research and the Arts in the framework of the Loewe Program (DynaMem). Research in the Stelzer lab is supported by the Deutsche Forschungsgemeinschaft (CEF-MC II, EXC-115). B.M. has additionally been supported by fellowships from the Joachim Herz Stiftung and Freunde und Förderer der Goethe-Universität. S.C.F. has been supported by an International Exchanges Grant from The Royal Society (IE141022).

REFERENCES

1. Rivera-Pérez, J. A., and A.-K. Hadjantonakis, 2015. The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harbor perspectives in biology* 7:a015867.

2. Downs, K. M., and T. Davies, 1993. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118:1255–1266.

3. Wilcox, A. J., C. R. Weinberg, J. F. O'connor, D. D. Baird, J. P. Schlatterer, R. E. Canfield, E. G. Armstrong, and B. C. Nisula, 1988. Incidence of early loss of pregnancy. *New England Journal of Medicine* 319:189–194.

4. Tarkowski, A. K., and J. Wróblewska, 1967. Development of blastomeres of mouse eggs isolated at the 4-and 8-cell stage. *Development* 18:155–180.

5. Johnson, M. H., and J. M. McConnell, 2004. Lineage allocation and cell polarity during mouse embryogenesis. *In* Seminars in cell & developmental biology. Elsevier, volume 15, 583–597.

6. Dziadek, M., 1979. Cell differentiation in isolated inner cell masses of mouse blastocysts in vitro: onset of specific gene expression. *Development* 53:367–379.

7. Chazaud, C., Y. Yamanaka, T. Pawson, and J. Rossant, 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell* 10:615–624.

8. Mihajlović, A. I., V. Thamodaran, and A. W. Bruce, 2015. The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. *Scientific reports* 5:15034.

9. Schrode, N., P. Xenopoulos, A. Piliszek, S. Frankenberg, B. Plusa, and A.-K. Hadjantonakis, 2013. Anatomy of a blastocyst: cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. *Genesis* 51:219–233.

10. Ohnishi, Y., W. Huber, A. Tsumura, M. Kang, P. Xenopoulos, K. Kurimoto, A. K. Oleś, M. J. Araúzo-Bravo, M. Saitou, A.-K. Hadjantonakis, et al., 2014. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nature cell biology* 16:27.

11. Plusa, B., A. Piliszek, S. Frankenberg, J. Artus, and A.-K. Hadjantonakis, 2008. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* 135:3081–3091.

12. Zernicka-Goetz, M., S. A. Morris, and A. W. Bruce, 2009. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nature Reviews Genetics* 10:467.

13. Singh, A. M., T. Hamazaki, K. E. Hankowski, and N. Terada, 2007. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells* 25:2534–2542.

14. Guo, G., M. Huss, G. Q. Tong, C. Wang, L. L. Sun, N. D. Clarke, and P. Robson, 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell* 18:675–685.

15. Frankenberg, S., F. Gerbe, S. Bessonnard, C. Belville, P. Pouchin, O. Bardot, and C. Chazaud, 2011. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Developmental cell* 21:1005–1013.

16. Kang, M., P. Xenopoulos, S. Muñoz-Descalzo, X. Lou, and A.-K. Hadjantonakis, 2013. Live imaging, identifying, and tracking single cells in complex populations in vivo and ex vivo. *In* Imaging and Tracking Stem Cells, Springer, 109–123.

17. Schrode, N., N. Saiz, S. Di Talia, and A.-K. Hadjantonakis, 2014. GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Developmental cell* 29:454–467.

18. Takaoka, K., and H. Hamada, 2012. Cell fate decisions and axis determination in the early mouse embryo. *Development* 139:3–14.

19. Yamanaka, Y., F. Lanner, and J. Rossant, 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* 137:715–724.

20. Schröter, C., P. Rué, J. P. Mackenzie, and A. M. Arias, 2015. FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development* 142:4205–4216.

21. Bessonnard, S., L. De Mot, D. Gonze, M. Barriol, C. Dennis, A. Goldbeter, G. Dupont, and C. Chazaud, 2014. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* 141:3637–3648.

22. De Mot, L., D. Gonze, S. Bessonnard, C. Chazaud, A. Goldbeter, and G. Dupont, 2016. Cell fate specification based on tristability in the inner cell mass of mouse blastocysts. *Biophysical journal* 110:710–722.

23. Tosenberger, A., D. Gonze, S. Bessonnard, M. Cohen-Tannoudji, C. Chazaud, and G. Dupont, 2017. A multiscale model of early cell lineage specification including cell division. *NPJ systems biology and applications* 3:16.

24. Rossant, J., C. Chazaud, and Y. Yamanaka, 2003. Lineage allocation and asymmetries in the early mouse embryo. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358:1341–1349.

25. Kurimoto, K., Y. Yabuta, Y. Ohinata, Y. Ono, K. D. Uno, R. G. Yamada, H. R. Ueda, and M. Saitou, 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic acids research* 34:e42–e42.

26. Krupinski, P., V. Chickarmane, and C. Peterson, 2011. Simulating the mammalian blastocyst-molecular and mechanical interactions pattern the embryo. *PLoS computational biology* 7:e1001128.

27. Gardner, R., M. Lyon, E. Evans, and M. Burtenshaw, 1985. Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. *Development* 88:349–363.

28. Horvitz, H. R., and I. Herskowitz, 1992. Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* 68:237–255.

29. Mathew, B., S. Muñoz-Descalzo, E. Corujo-Simon, C. Schröter, E. H. Stelzer, and S. C. Fischer, 2019. Mouse ICM organoids reveal three-dimensional cell fate clustering. *Biophysical journal* 116:127–141.

30. Karolak, A., D. A. Markov, L. J. McCawley, and K. A. Rejniak, 2018. Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues. *Journal of The Royal Society Interface* 15:20170703.

31. Drasdo, D., and S. Höhme, 2005. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Physical biology* 2:133.

32. Drasdo, D., S. Hoehme, and M. Block, 2007. On the role of physics in the growth and pattern formation of multi-cellular systems: what can we learn from individual-cell based models? *Journal of Statistical Physics* 128:287.

33. Germann, P., M. Marin-Riera, and J. Sharpe, 2019. ya||a: GPU-Powered Spheroid Models for Mesenchyme and Epithelium. *Cell Systems* 8:261 – 266.

34. Rejniak, K. A., and A. R. Anderson, 2011. Hybrid models of tumor growth. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 3:115–125.

35. Stichel, D., A. M. Middleton, B. F. Müller, S. Depner, U. Klingmüller, K. Breuhahn, and F. Matthäus, 2017. An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. *NPJ systems biology and applications* 3:5.

36. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain, 2002. Stochastic gene expression in a single cell. *Science* 297:1183–1186.

37. Dietrich, J.-E., and T. Hiiragi, 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development* 134:4219–4231.

38. Simon, C. S., A.-K. Hadjantonakis, and C. Schröter, 2018. Making lineage decisions with biological noise: Lessons from the early mouse embryo. *Wiley Interdisciplinary Reviews: Developmental Biology* 7:e319.

39. Streichan, S. J., C. R. Hoerner, T. Schneidt, D. Holzer, and L. Hufnagel, 2014. Spatial constraints control cell proliferation in tissues. *Proceedings of the National Academy of Sciences* 111:5586–5591.

40. Artus, J., and C. Chazaud, 2014. A close look at the mammalian blastocyst: epiblast and primitive endoderm formation. *Cellular and molecular life sciences* 71:3327–3338.

41. Krens, S. G., and C.-P. Heisenberg, 2011. Cell sorting in development. *In* Current topics in developmental biology, Elsevier, volume 95, 189–213.

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

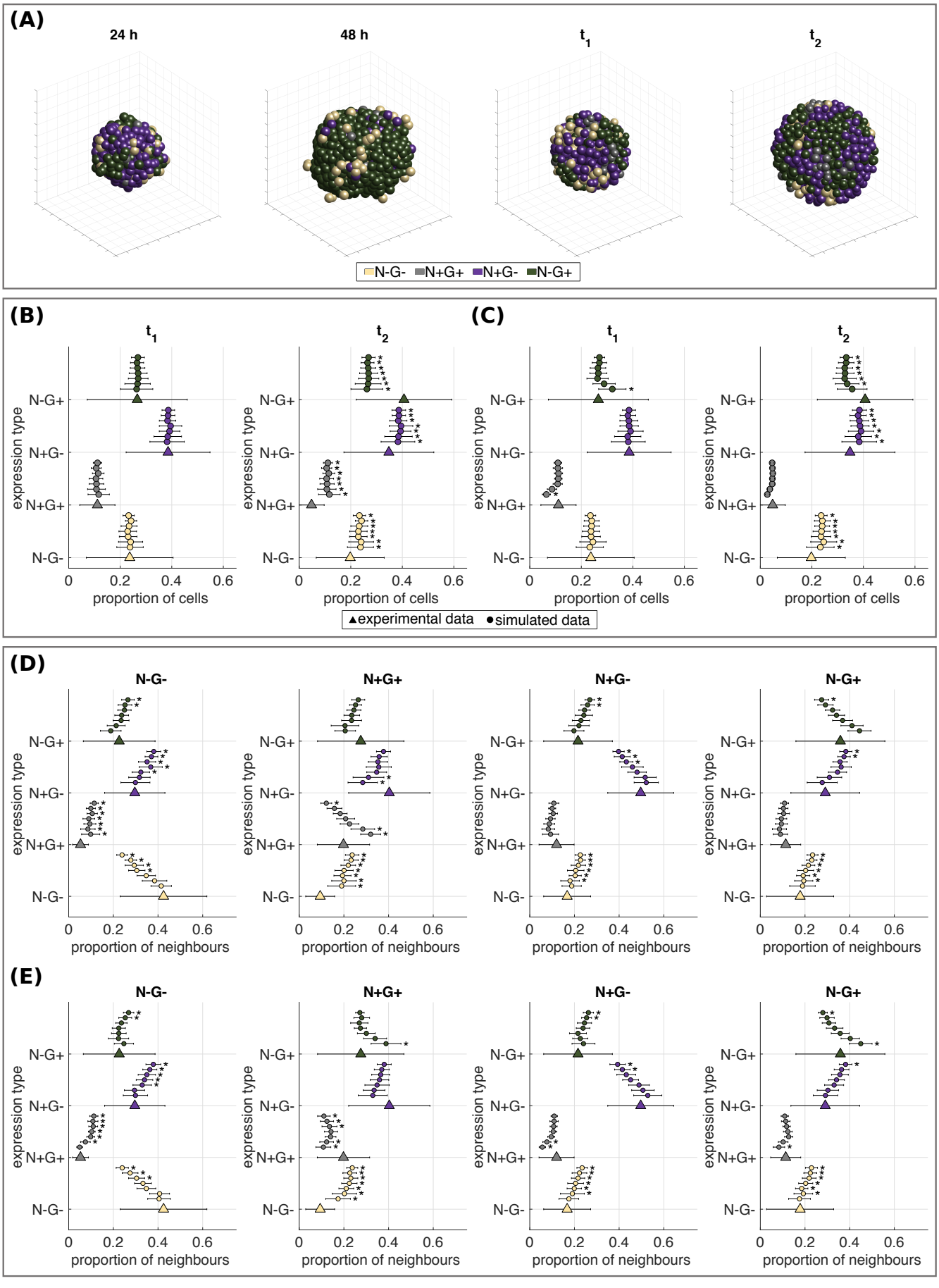


Figure 1

Figure 1: Expression type composition of organoids. (A) ICM organoids (experimental data) for 24 h and 48 h and simulated ICM spheroids for t_1 and t_2. (B, C) Expression type composition of ICM organoids and ICM spheroids shown as percentage of the total number of cells within ICM organoids at t_1 and t_2. Statistically significant differences between the cell fate proportion of ICM organoids and ICM spheroids are indicated by stars ($p < 0.05$; using a Wilcoxon-Mann-Whitney test with Bonferroni correction). Simulations were performed under the assumption H1 (B) and the assumption H2 (C). (D, E) Expression type composition of neighbouring cells shown as percentage of the total of neighbouring cells at t_1. Simulations were performed under the assumption H1 (D) and the assumption H2 (E). Experimental data from Mathew *et al.* (2019) are indicated by triangles. Simulation results for different t_0 are indicated by circles. The error bars indicate the standard deviation. t_0 from lowest line to top: 75, 100, 150, 200, 250, 300 and 400 cells. Statistically significant differences between the neighbourhood structure of 24 h old ICM organoids and ICM spheroid patterns are indicated by stars ($p < 0.05$; using a Wilcoxon-Mann-Whitney test with Bonferroni correction).

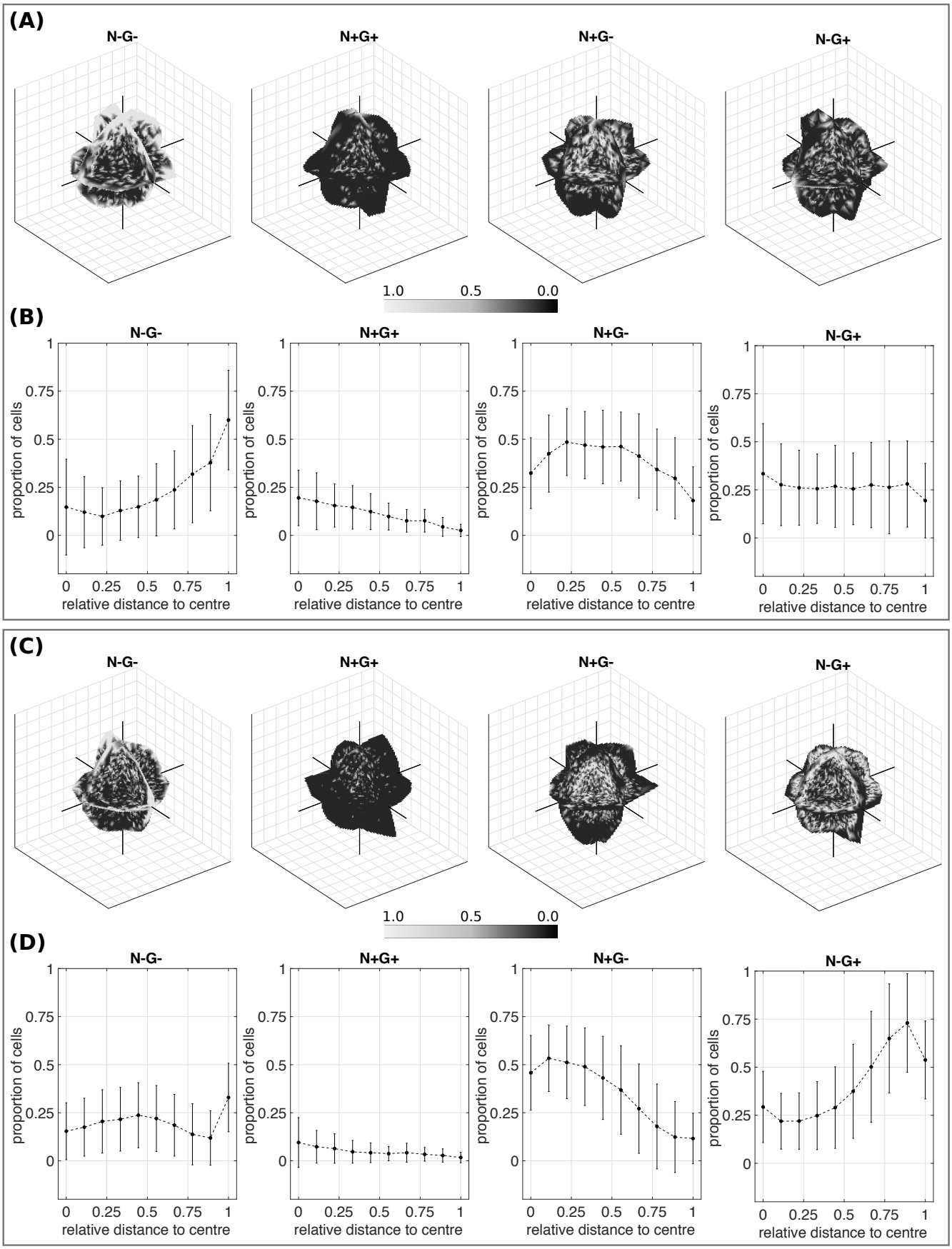


Figure 2

Figure 2: Expression type cluster analysis for ICM organoids. (A) 24 h old ICM organoids; (C) 48 h old ICM organoids. Black indicates the absence of an expression type, white indicates the presence of the selected expression type, respectively. Shown are slices through the ICM organoids at $x = 0$, $y = 0$ and $z = 0$. (B, D) Expression type compositions in dependence of the relative distance to the ICM organoid centre for 24 h old ICM organoids (B) and for 48 h old ICM organoids (D). Cells are sorted according to their distance to the ICM organoid centre of mass and binned into 10 groups. Points indicate the mean proportion of a cell fate type for the 10 bins, the bars denote the standard deviation.

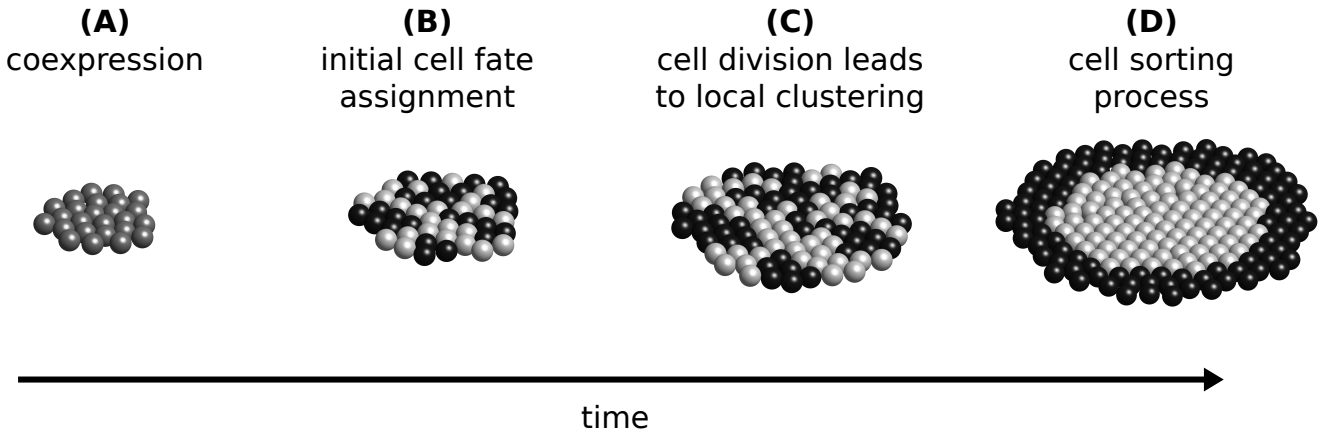
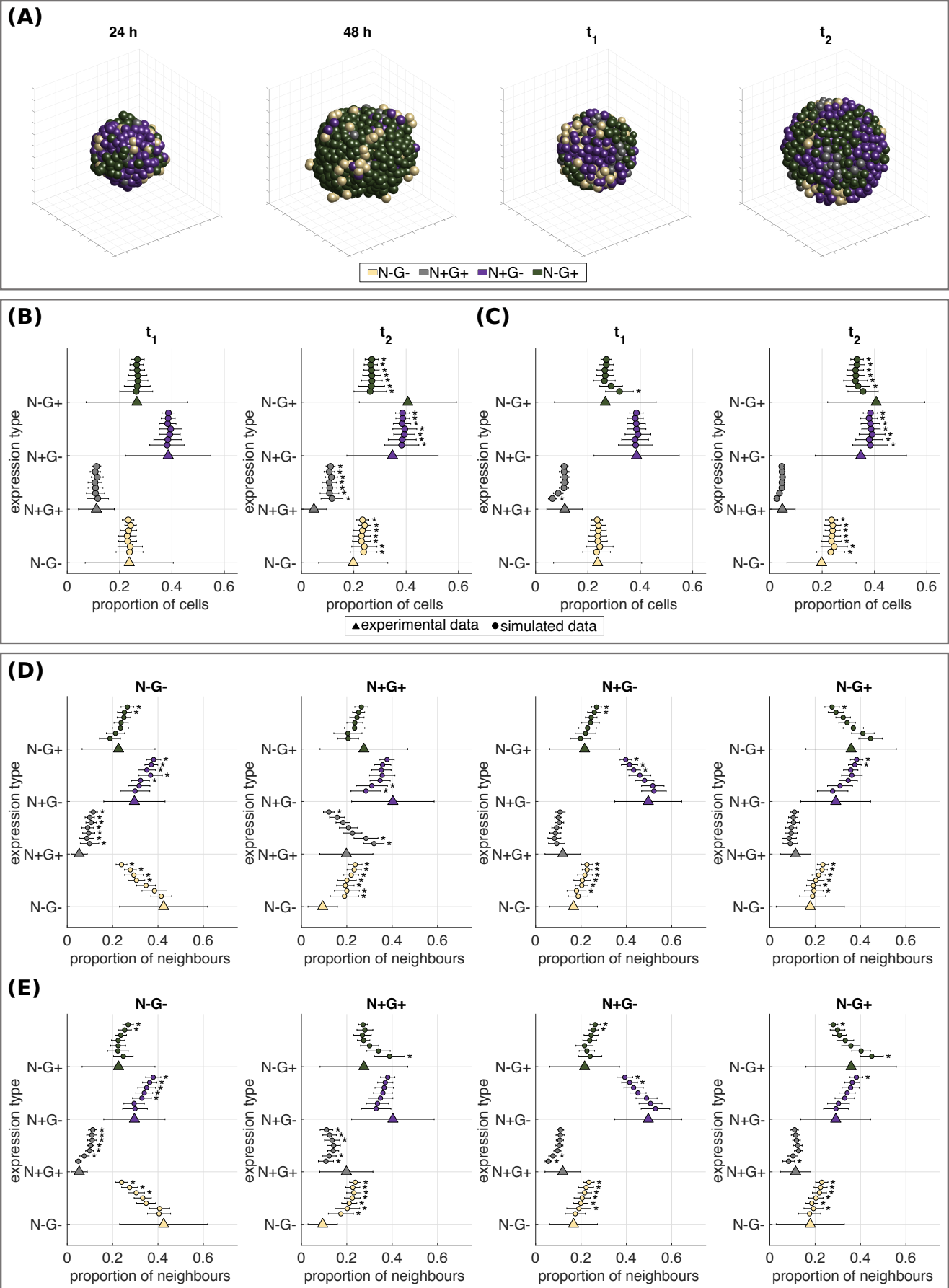
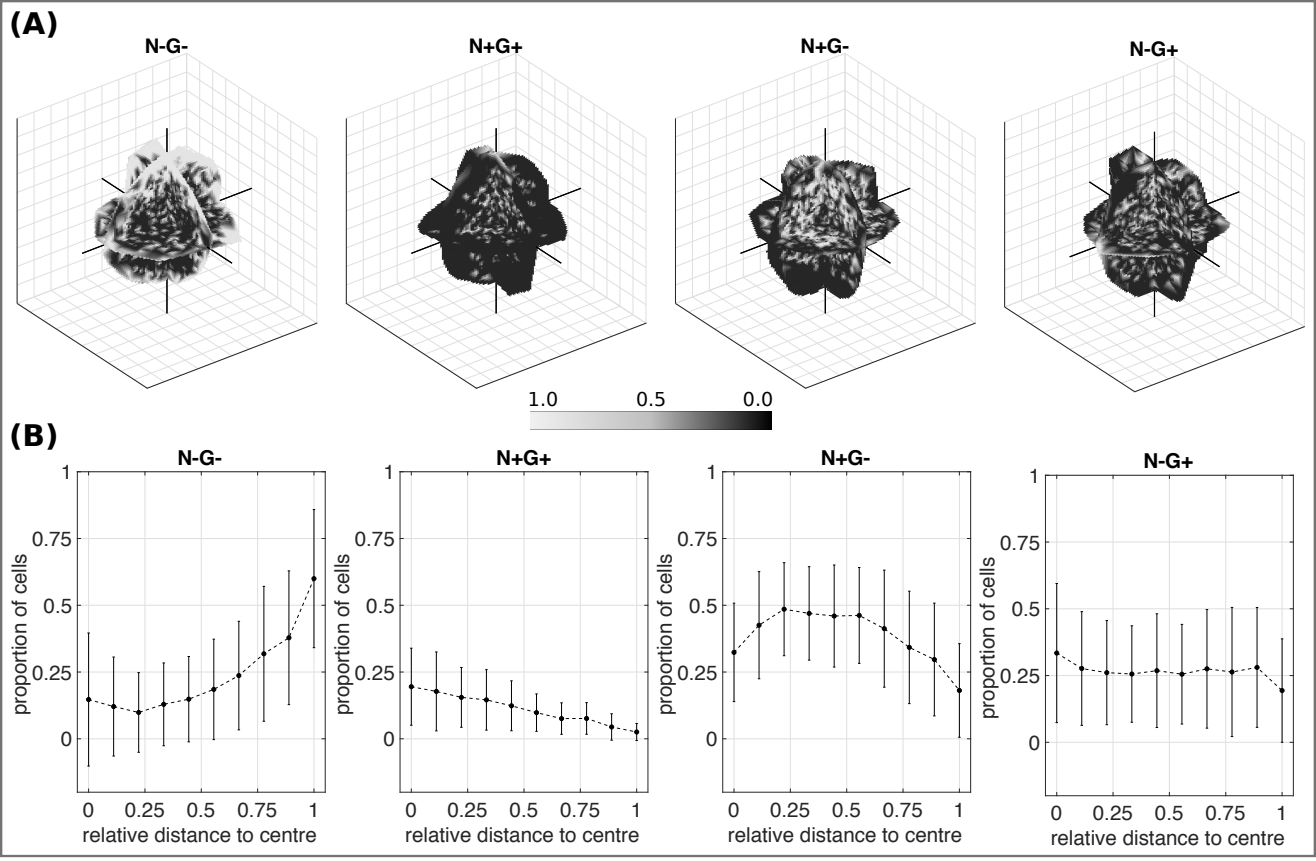
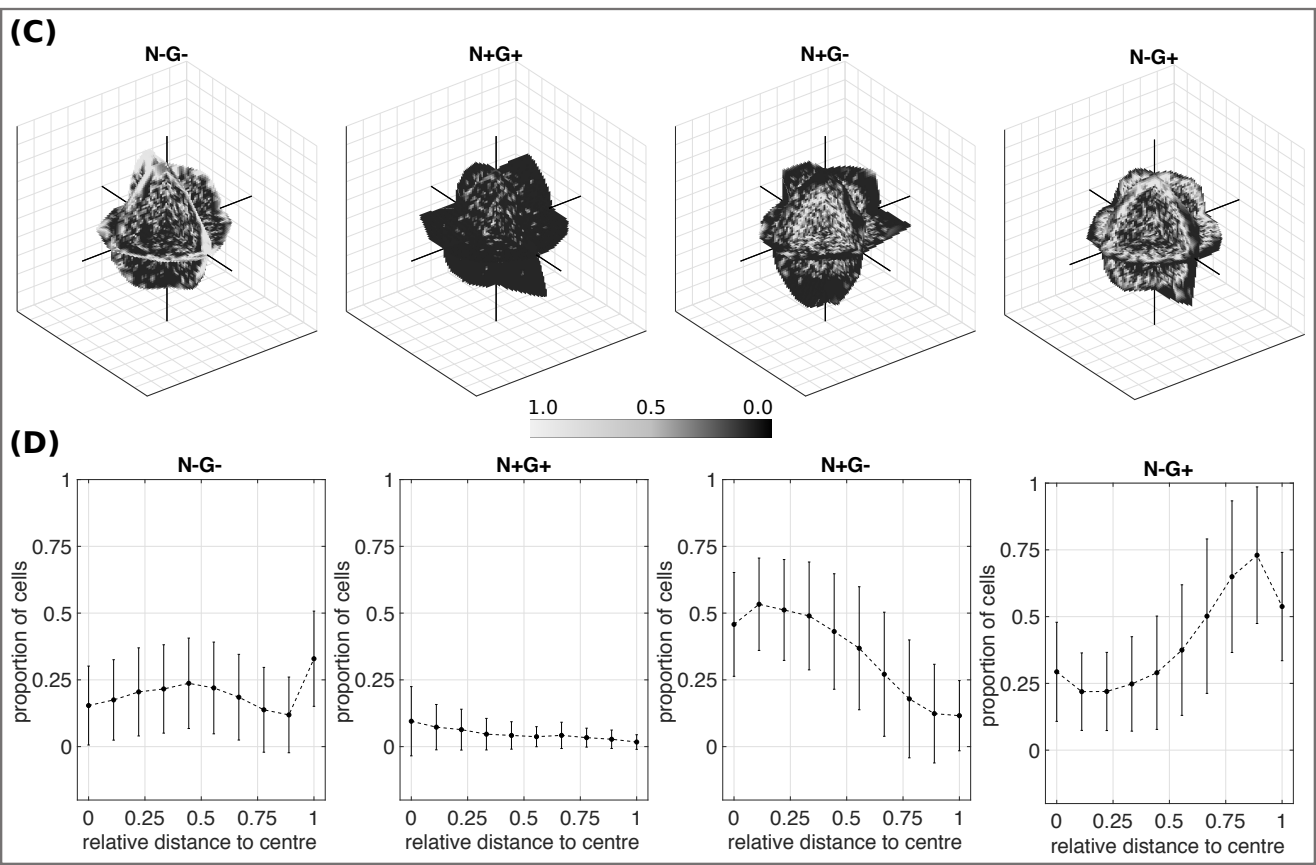
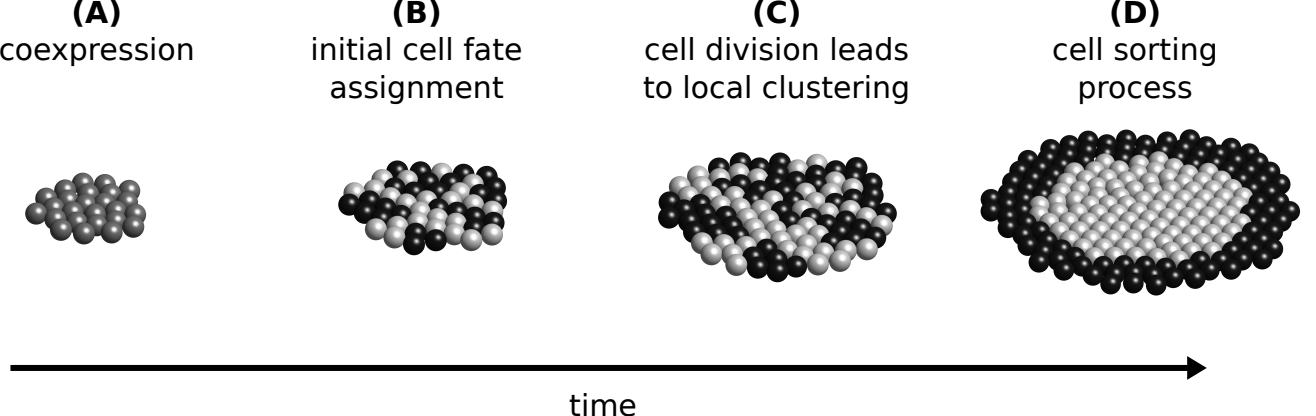


Figure 3: Working model. Over time four phases lead to separation of GATA6 positive (black) and NANOG positive (white) cells. In the beginning all cells coexpress both cell fate markers (A), followed by a random cell fate assignment (B). Cell division leads to a local clustering of cells sharing the same cell fate (C) and eventually a sorting mechanism leads to a phase separation of GATA6 positive and NANOG positive cells.


(A)
24 h
48 h
t1
t2
N-G-
N+G+
N+G-
N-G+
(B)
t1
t2
expression type
N-G+
N+G-
N+G+
N-G-
0
0.2
0.4
0.6
proportion of cells
(C)
t1
t2
expression type
N-G+
N+G-
N+G+
N-G-
0
0.2
0.4
0.6
proportion of cells
▲experimental data ●simulated data
(D)
N-G-
N+G+
N+G-
N-G+
expression type
N-G+
N+G-
N+G+
N-G-
0
0.2
0.4
0.6
proportion of neighbours
(E)
N-G-
N+G+
N+G-
N-G+
expression type
N-G+
N+G-
N+G+
N-G-
0
0.2
0.4
0.6
proportion of neighbours


(A)
N-G-
N+G+
N+G-
N-G+
1.0
0.5
0.0
(B)
N-G-
N+G+
N+G-
N-G+
proportion of cells
relative distance to centre
0
0.25
0.5
0.75
1


(C)
N-G-
N+G+
N+G-
N-G+
1.0
0.5
0.0
(D)
N-G-
N+G+
N+G-
N-G+
proportion of cells
relative distance to centre
0
0.25
0.5
0.75
1


(A)
coexpression
(B)
initial cell fate
assignment
(C)
cell division leads
to local clustering
(D)
cell sorting
process
time