Cell-cell communication through FGF4 generates and maintains robust proportions of differentiated cell types in embryonic stem cells

Dhruv Raina[1§], Azra Bahadori[1,2§], Angel Stanoev[1,3], Michelle Protzek[1], Aneta Koseska[1,4], Christian Schröter[1,*]

[1] Department of Systemic Cell Biology, Max Planck Institute of Molecular Physiology, Dortmund, Germany

[2] Current address: Center for Chromosome Stability, University of Copenhagen, Copenhagen, Denmark

[3] Current address: Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

[4] Current address: Cellular computations and learning, research center caesar, Bonn, Germany

§ These authors contributed equally

[*] For correspondence: christian.schroeter@mpi-dortmund.mpg.de

Key words

FGF signalling, embryonic stem cells, preimplantation development, primitive endoderm, epiblast, dynamical system

Summary statement

Experiments in embryonic stem cells reveal how the coupling of cell fate decisions in a population leads to the self-organized differentiation of stable ratios of cell types.

Abstract

During embryonic development and tissue homeostasis, reproducible proportions of differentiated cell types are specified from populations of multipotent precursor cells. Molecular mechanisms that enable both robust cell type proportioning despite variable initial conditions in the precursor cells, as well as the re-establishment of these proportions upon

© 2021. Published by The Company of Biologists Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Development • Accepted manuscript

perturbations in a developing tissue remain to be characterised. Here we report that the differentiation of robust proportions of epiblast-like and primitive endoderm-like cells in mouse embryonic stem cell cultures emerges at the population level through cell-cell communication via a short-range FGF4 signal. We characterise the molecular and dynamical properties of the communication mechanism, and show how it controls both robust cell type proportioning from a wide range of experimentally controlled initial conditions, as well as the autonomous re-establishment of these proportions following the isolation of one cell type. The generation and maintenance of reproducible proportions of discrete cell types is a new function for FGF signalling that may operate in a range of developing tissues.

Introduction

The differentiation of specialised cell types from populations of multipotent precursor cells is the basis of embryonic development and tissue homeostasis in the adult. Developing tissues generally produce and maintain a standard end result consisting of defined proportions of differentiated cell types despite biological noise and perturbations, a behaviour termed canalization (Waddington, 1942). Molecular mechanisms that contribute to canalized development by controlling the proportions of differentiated cell types remain to be characterized.

Mammalian preimplantation development is a prime example for developmental canalization. The size of the three lineages trophoectoderm (TE), epiblast (Epi), and primitive endoderm (PrE) is remarkably constant between mouse preimplantation embryos (Saiz et al., 2016). Furthermore, mammalian embryos can regulate the proportions of the three lineages following splitting, fusing, or the addition of embryonic stem cells (ESCs), such that the embryo is capable of post-implantation development (Arias et al., 2013; Bedzhov et al., 2014; Bradley et al., 1984; Gardner, 1968; Tarkowski, 1959; Tarkowski, 1961). The differentiation of Epi and PrE cells from inner cell mass (ICM) cells is controlled by transcription factors such as NANOG and GATA6, which mark and specify Epi and PrE cells, respectively. These factors are initially co-expressed in ICM cells, and become mutually exclusive as cells differentiate (Chazaud et al., 2006; Plusa et al., 2008; Simon et al., 2018). In addition, PrE differentiation requires fibroblast growth factor (FGF)/ extracellular regulated kinase (ERK) signalling (Chazaud et al., 2006; Kang et al., 2017; Krawchuk et al., 2013; Molotkov et al., 2017). Current models for cell differentiation in the ICM posit that mutually repressive interactions between transcription factors together with heterogeneous FGF/ERK signalling allocate individual cells to one of the two lineages (Bessonnard et al., 2014; Caluwé et al.,

Development • Accepted manuscript

2019; Chickarmane and Peterson, 2008; Mot et al., 2016; Yamanaka et al., 2010). More recently, FGF signalling has been shown to regulate Epi and PrE lineage sizes following the addition or ablation of cells, suggesting that it orchestrates cell differentiation at the population level (Saiz et al., 2020).

Population-level mechanisms for cell differentiation are an attractive solution to the problem of developmental canalization, as it has been shown theoretically that populations of communicating cells can re-establish specific cell type proportions following perturbations (Stanoev et al., 2021). This theory furthermore predicts that the differentiation outcomes are insensitive to initial conditions, since the heterogeneous cell identities represent a collective state that is generated and maintained on the level of the communicating population, rather than being specified intrinsically in each cell. Identifying a molecular mechanism that leads to such emergent phenomena however requires an experimental system in which both the initial conditions in the precursor population as well as cell-cell communication can be precisely controlled.

The specification of Epi- and PrE-like cells from ESCs following the transient expression of exogenous GATA factors is a suitable model system to investigate mechanisms of cell type proportioning. In serum-containing medium, the specification of PrE-like cells requires above-threshold levels of inducible GATA factors and ERK activity, consistent with a model in which mutually repressive interactions between NANOG and GATA factors together with FGF/ERK signalling control differentiation at the single cell level (Fig. 1A, Schröter et al., 2015). To explore the role of population-level mechanisms, we here study cell differentiation in defined serum-free media where signalling cues are produced solely by the cells themselves. Under these conditions, robust proportions of Epi- and PrE-like cells differentiate from a wide range of initial conditions generated by experimentally controlled GATA expression levels, and regenerate from populations of purified PrE-like cells. We use communication-deficient mutant cell lines and simulations to demonstrate that cell-cell communication via a short-range FGF4 signal is the minimal molecular mechanism underlying this robust population-level behaviour. These results provide evidence that local cell-cell communication via a secreted growth factor can contribute to the differentiation of reproducible global proportions of specialised cell types.

Development • Accepted manuscript

Results

Differentiation of robust proportions of Epi- and PrE-like cells in ESC cultures

To study population-level mechanisms that control the proportions of PrE-like and Epi-like cells in ESC populations, we used ESC lines carrying doxycycline-inducible GATA4-mCherry constructs that were integrated via PiggyBac transgenesis and that allowed sampling a wide range of GATA4-mCherry expression levels. Cells were kept in chemically defined minimal N2B27 medium supplemented with the MEK inhibitor PD0325901 (PD03), the GSK3 inhibitor CHIR99021, and the cytokine LIF (2i + LIF medium, (Ying et al., 2008)) during routine culture and transgene induction to maintain pluripotency. To initiate differentiation, cells were switched to N2B27 only by simultaneously removing doxycycline, LIF, and inhibitors (Fig. 1B).

Under these conditions, an 8 h pulse of GATA4-mCherry expression triggered the rapid downregulation of NANOG and the upregulation of endogenous GATA6 (Fig. S1A,B). While some cells co-expressed NANOG and GATA6 8 h after the end of the GATA4-mCherry pulse, mutually exclusive expression of GATA6 and NANOG was established within 16 h and became more pronounced until 40 h (Fig. S1A-C). From this point onwards, the proportion of GATA6-positive cells decreased, and the proportion of GATA6-negative increased (Fig. S1D). The separation of the two cell populations in the NANOG;GATA6 expression space suggests that this shift in proportions beyond 40 h is a consequence of differential proliferation rates rather than fate transitions in individual cells. NANOG-expression levels in the GATA6-negative cell cluster decreased over time, consistent with differentiation along the embryonic epiblast lineage (Fig. S1C). At 40 h, expression of the three PrE-markers GATA6, SOX17 and Laminin was mutually exclusive with NANOG expression (Niakan et al., 2010, Fig. S1 E, F). Thus, transient expression of GATA4-mCherry followed by 40 h of differentiation in defined, growth-factor-free medium subdivides an initially homogeneous culture into two cell types with Epi- and PrE-like characteristics.

To determine how the proportions of the two cell types were affected by expression levels of the inducible GATA4-mCherry protein, we titrated GATA4-mCherry expression by staggering the initiation of doxycycline induction in time, while keeping the duration of the entire experiment constant (Fig. 1B). GATA4-mCherry levels increased, and NANOG expression levels decreased with longer doxycycline induction time, both in the population as well as in individual cells (Figs 1C, D, S1A and S2A). However, despite these different transcription factor expression levels at the start of differentiation, we observed similar proportions of both GATA6+;NANOG- PrE-like and GATA6-;NANOG+ Epi-like cells 40 h

Development • Accepted manuscript

later (Figs 1D and S2A). The proportion of PrE-like cells slightly increased from 43.6 ± 12.1% for 1 h to 59.7 ± 11.9% ($p = 0.02$) and 57.2 ± 8.7% ($p = 0.04$) for 4 h and 8 h of induction, respectively, but did not show any differences between all other conditions ($p > 0.05$, Tukey's multiple comparisons test). The proportion of Epi-like cells was stable for different induction times ($p > 0.05$, Figs 1E and S2B). Thus, in minimal medium, a wide range of inducible GATA4-mCherry expression levels leads to similar proportions of differentiated cell types (Fig. 1F). This robust proportioning of cell types could be a population-level phenomenon, or alternatively have a cell-intrinsic basis such as the pre-specification of cell types or a limited differentiation potential in a subset of cells. To rule out this latter possibility, we promoted ERK activity by adding recombinant FGF4 during the differentiation phase which reveals the differentiation potential of single cells. In the presence of FGF4, the proportion of PrE-like cells was higher than upon differentiation in N2B27 alone, and significantly increased with doxycycline induction time between most of the conditions ($p < 0.05$, except for 1 h vs. 2 h and 4 h vs. 8 h, Fig. S2C). Thus, in the presence of exogenous signals GATA4-mCherry levels control cell type proportions, in line with previous findings (Schröter et al., 2015).

To separate the effects of extended GATA4-mCherry expression from dosage effects, we analysed cell differentiation in four clonal cell lines with independent integrations of the GATA4-mCherry transgene. GATA4-mCherry expression levels following 8 h of doxycycline induction varied widely between these lines, both at the single cell and the population level (Figs S3A and S3B). Yet, in three out of the four clones, similar proportions of PrE-like cells differentiated upon doxycycline removal and culture in N2B27 ($p < 0.05$, except for C2 vs. all other clones, Figs S3C,D). When we added recombinant FGF4 during the differentiation phase, the fraction of PrE-like cells was higher than upon differentiation in N2B27 alone, and systematically increased with GATA4-mCherry induction levels ($p < 0.005$, ANOVA test for linear trend). In the clonal line with the highest GATA4-mCherry levels, the proportion of PrE-like cells reached a maximum of 98.8 ± 2.0% upon addition of exogenous FGF4 (Fig. S3D), demonstrating that almost all cells have PrE-like differentiation potential following sufficiently strong GATA4-mCherry expression. Taken together, these results indicate that in minimal medium, the robust proportioning of cell types is established at the population level through cell-cell signalling.

Development • Accepted manuscript

Differentiating ESCs communicate via FGF4

To identify candidate mechanisms for cell-cell communication that could underlie this robust cell type proportioning we focused on FGF4, since FGF/ERK signalling is required for PrE differentiation both in ESCs and in the embryo (Kang et al., 2012; Krawchuk et al., 2013; Schröter et al., 2015), and paracrine FGF4 is the main activator of ERK in ESCs (Kunath et al., 2007).

To investigate how GATA4-mCherry induction levels affect FGF4 signalling, we integrated a *Sprouty4*$^{H2B\text{-}Venus}$ transcriptional reporter as a quantitative readout for long-term FGF4 signalling (Morgani et al., 2018) in the inducible cell lines. Longer doxycycline induction times corresponding to higher GATA4-mCherry expression levels resulted in reduced mean reporter fluorescence after 24 h of differentiation (Fig. 2A). To test whether this negative correlation between GATA4-mCherry levels and FGF4 signalling was caused by direct regulation of *Fgf4* transcription through GATA factors or via indirect regulation through NANOG (Frankenberg et al., 2011), we used *in situ* mRNA staining for *Fgf4* to determine its expression dynamics following GATA4-mCherry induction. *Fgf4* mRNA was strongly expressed in most cells before induction in 2i medium, but its levels started to decline within 2 h after the start of doxycycline induction (Fig. 2B, left). After 8 h, *Fgf4* mRNA levels were strongly reduced, except in cells with low GATA4-mCherry expression levels (arrowheads in Fig. 2B). As most cells still express NANOG at this time (Fig. S1A), the rapid downregulation of *Fgf4* mRNA suggests a direct transcriptional regulation through GATA4-mCherry. After 40 h of differentiation in N2B27, *Fgf4* mRNA expression was mutually exclusive with *Gata6* mRNA in cell cultures that had received an 8 h doxycycline pulse (Fig. 2C). Without a prior doxycycline pulse, *Fgf4* mRNA continued to be expressed in the majority of cells after 40 h of culture in N2B27 (Fig. 2C), although NANOG protein was almost completely downregulated (Fig. 2D). Taken together, these data indicate that GATA factors directly regulate *Fgf4* transcription in ESCs. This possibility was further supported by the presence of a GATA6 binding peak approx. 10 kb upstream of the *Fgf4* start codon in a published ChIPseq dataset (Wamaitha et al., 2015) that contains a large number of GATA consensus binding sites (Fig. S4A,B). However, in cells where this putative GATA6 binding site had been deleted, *Fgf4* mRNA expression was downregulated to levels similar to those observed in wild type cells, both at the end of an 8 h doxycycline pulse, and after 40 h of differentiation (Fig. S4 C-E). This suggests that *Fgf4* regulation by GATA factors occurs through multiple, possibly redundant gene regulatory elements.

Development • Accepted manuscript

Having shown how the cell-intrinsic transcriptional circuits affect FGF4 signalling, we next tested how these circuits were affected by FGF4 dose. To be able to control FGF4 levels, we mutated the *Fgf4* gene in the GATA4-mCherry inducible cells. In line with previous reports (Kunath et al., 2007), *Fgf4* mutant cells continued to express high levels of NANOG upon culture in N2B27, indicative of a failure to initiate epiblast differentiation. This phenotype could be rescued to wild-type levels by addition of recombinant FGF4 (Fig. S5A,B). When pulsed for 8 h with doxycycline before culture in N2B27, *Fgf4* mutant cells likewise continued to express high levels of NANOG, and showed almost no signs of PrE-like differentiation, in contrast to the wild type control (Fig. 2E,F). PrE-like differentiation could be rescued by supplementing recombinant FGF4 during the differentiation phase, resulting in two discrete cell types that had similar NANOG- and GATA6-expression profiles as differentiating wild type cells (Fig. S5C). The proportions of these cell types depended on FGF4 concentration (Figs 2E,F, S5C). Thus, FGF4 signalling and the cell-intrinsic transcriptional circuits underlying cell differentiation mutually regulate each other in a dose-dependent fashion. Communication via FGF4 is therefore a potential mechanism for coordinating cell differentiation in the population.

Paracrine FGF4 in ESCs acts locally

We next sought to determine the spatial range of FGF4 signalling in ESCs. We first tested the role of global communication by comparing differentiation outcomes at different medium-to-cell ratios (Fig. 3A). If FGF4 acted globally, ligand concentration would equilibrate in the medium, such that larger volumes would effectively reduce FGF4 concentration, and decrease the proportion of PrE-like cells. In contrast to this expectation, cell type proportions changed negligibly with media volume (Fig. 3A), indicating that dilution of FGF4 ligands in the medium does not strongly affect cell type proportioning.

To test whether in contrast the communication is governed by local FGF4 signalling, we disrupted cell-cell contacts by replating cells at different densities immediately after doxycycline induction (Fig. 3B). Replating strongly reduced the proportion of PrE-like cells compared to the non-trypsinised control ($p < 0.05$, Dunnett's multiple comparison test, compare second to three rightmost columns in Fig. 3B), indicating that cell-cell contacts in intact colonies support PrE-like differentiation. Furthermore, the proportion of PrE-like cells systematically increased with cell density (three rightmost columns in Fig. 3B, $p < 0.05$, one-way ANOVA test for linear trend). These data suggest that cell-cell communication via FGF4 occurs locally and is positively influenced by cell-cell contacts.

Development • Accepted manuscript

To directly measure the spatial range of FGF4 signalling in ESC colonies, we seeded single labelled wild type cells onto a layer of *Fgf4* mutant *Spry4*$^{H2B\text{-}Venus}$ reporter cells (Morgani et al., 2018). 12 h after the addition of wild type cells, H2B-Venus was strongly expressed in a halo of reporter cells immediately surrounding the signal-emitting cells, but reporter expression dropped precipitously further away from the sender cells (Fig. 3C). The spatial profile of the H2B-Venus signal was well approximated by a plateau of ~ 14.4 µm, followed by an exponential decay with a decay length of ~ 11 µm (Fig. 3D). This is likely an overestimate of the immediate effective range of paracrine FGF4 signalling, as the transcriptional reporter integrates signalling activity over the entire duration of the experiment, during which cell divisions and movement will increase the distance between signal-sending and –receiving cells. Delaunay triangulation revealed that 16 h after the initiation of differentiation, the mean distance between a cell and its nearest and second-nearest neighbours was 14.0 ± 3.2 µm and 25.5 ± 5.3 µm, respectively (Methods, Fig. S6). Thus, the range of cell-cell communication via FGF4 is spatially restricted and mainly couples nearest and second-nearest neighbours.

We further confirmed the spatially restricted activity of FGF4 using cell differentiation as a read-out. When we added a low number of labelled wild type cells to a culture of *Fgf4* mutant GATA4-mCherry inducible cells immediately after the end of a doxycycline pulse, PrE-like cells were almost exclusively found in colonies containing *Fgf4* wild type cells and often localized close to the *Fgf4* sender cells 16 h later (Fig. 3 E,F).

Cell-cell communication via FGF4 underlies robustness of cell type proportions

We next asked if local communication via FGF4 was the molecular mechanism underlying cell type proportioning. We titrated GATA4-mCherry expression levels by varying induction time, and compared the robustness of cell type proportions between wild type cells that can sense and secrete FGF4, and communication-deficient *Fgf4* mutant cells rescued with a fixed dose of recombinant FGF4 (Fig. 4A-C). While the proportions of Epi- and PrE-like cells remained relatively constant in the wild type (Fig. 4A-C, left column, $p > 0.05$, except for the proportion of PrE-like cells for 1 h induction vs. all other conditions), they were strongly dependent on induction times in *Fgf4* mutant cells rescued with 10 ng/ml FGF4 (right column, $p < 0.05$, except for the proportion of Epi-like cells for 1 h vs. 2 h and 4 h vs. 8 h induction). Rescuing *Fgf4* mutant cells with lower doses of FGF4 decreased the proportion of PrE-like cells and increased the proportion of Epi-like cells, but for all FGF4 doses tested, cell type proportions changed with induction time (Fig. S7A). Thus, cell type proportions in

Development • Accepted manuscript

rescued *Fgf4* mutant cells strongly depend on initial transcription factor expression levels, in contrast to wild type cells, which can buffer cell type proportions over a wide range of starting conditions. The critical difference between *Fgf4* wild type and rescued mutant cultures is the connectivity of the cellular network: While cell differentiation in wild type cultures is coupled via FGF4, cells in *Fgf4* mutant cultures take differentiation decisions largely autonomously. We conclude that communication via FGF4 is the molecular mechanism responsible for robust cell type proportioning in the population.

To further explore how this population-level behaviour of robust proportioning arises from cell-cell communication, we determined the dynamical properties of a communicating cell population with numerical simulations. We developed a mathematical description by considering a previously characterized circuit consisting of mutually repressive interactions between GATA factors and NANOG (Schröter et al., 2015) that communicates among single cells through FGF signalling (Fig. 4D, left). Specifically, we posit that GATA factors repress FGF4 expression (Fig. 2), and that FGF signalling represses Nanog expression (Hamilton and Brickman, 2014; Schröter et al., 2015). Cell-cell communication was set between nearest and second-nearest neighbours in a population of N = 10000 cells on a 100 x 100 square grid.

In the simulations, we considered a range of initial conditions, from all cells being GATA-positive initially to all cells being NANOG-positive initially, to mimic the experimental settings (Fig. 4E, left column). Within the coupled population, a stable proportion of two distinct gene expression patterns - (NANOG+, GATA-) or (NANOG-, GATA+) - was established (Fig. 4E-G, middle) from all starting conditions. To demonstrate that this robustness of cell type proportions is a consequence of coupling via FGF, we compared the results to a model where we replaced communication with a constant exogenous FGF input. This configuration mirrors the situation in the rescued *Fgf4* mutant and effectively models the single cell behaviour where cell differentiation is exclusively governed by the dynamics of the mutually repressive NANOG-GATA circuit. In this non-communicating model, cell type proportions strongly depended both on the initial conditions distributions (Figs 4D-G, right), as well as on FGF4 input strength (Fig. S7B), recapitulating experimental results in the rescued *Fgf4* mutant (Figs 4A-C, S7A). Finally, we tested the effect of adding exogenous signal to coupled wild type cells in order to override cell-cell communication. Both in experiments and simulations, this increased the proportion of (NANOG-, GATA+) cells at the expense of (NANOG+, GATA-) cells (Figs S2B, S7C,D). Taken together, this congruence between the theoretical and experimental results indicates that recursive cell-cell

Development • Accepted manuscript

communication via FGF signalling is sufficient to recapitulate our observation of robust cell type proportioning.

To explore the dynamical basis of robust cell type proportioning, we performed bifurcation analysis of a minimal system of N = 2 communicating cells. This revealed the presence of an inhomogeneous steady state (IHSS), a new collective dynamical state in the coupled system (Fig. S8). An IHSS is composed of mutually exclusive gene expression patterns, it is generated and dynamically maintained via communication at the population level, and has recently been proposed as a generic mechanism for robust cell type proportioning (Stanoev et al., 2021). The emergence of an IHSS is thus a possible dynamical basis for robust cell type proportioning in differentiating ESCs.

Spatial organization of cell types indicates a shift from FGF4-dependent to FGF4-independent patterning mechanisms

The short spatial range of FGF4 signals in ESC cultures suggests that communication via FGF4 not only leads to robust global cell type proportions, but that the differentiated cell types should be also arranged in spatial patterns with local structure. We therefore analysed the spatial arrangement of cell types at different time points after the initiation of differentiation. Staining for GATA6 and NANOG indicated that cell types were well mixed after 16 h and 24 h of differentiation, but clustering, particularly of GATA6-; NANOG+ Epi-like cells, was observed after 40 h (Fig. 5A). For further analysis, we focused exclusively on GATA6+; NANOG- (G+) and GATA6-; NANOG+ (N+) cells that we identified with a gaussian mixture model (Methods). Analysis of the cell type composition of the immediate neighbourhood N+ and G+ cells corroborated a transition from well-mixed to clustered patterns (Fig. S9A).

To quantify cluster sizes during the differentiation time course, we computed the scaled fraction of N+ cells in neighbourhoods of increasing radius around all N+ cells in a field of view (Fig. 5B, Methods). The value of the scaled fraction is 1 as long as all cells in the neighbourhood are N+, and approaches zero when the composition of the local neighbourhood equals the global composition of cell types. We defined the cluster radius as the distance around N+ cells at which this scaled fraction drops to 0.5 (dashed lines in Fig. 5B). The measure of cluster radius allows quantifying the length scales of spatially irregular features, however, its absolute value is smaller than the physical size of the spatial features. In wild type cells, the median cluster radius was 13.0 µm and 14.7 µm at 16 h and 24 h of

Development • Accepted manuscript

differentiation, respectively, and increased to 29.0 µm at 40 h of differentiation, corresponding to approximately 0.9, 1.0, and 2.1 cell diameters (Fig. 5C, orange).

To investigate whether these experimentally determined wild type cluster patterns are consistent with short range signalling, we quantified the cell type spatial organization in model simulations. We modelled cluster formation due to cell division by considering a dividing cell population, where the grid size was doubled at each cell division event, and the daughter cells inherited gene expression states from the mother cell (Fig. S10A). In the wild type case, the cluster radius increased only slightly, from 0.6 to 0.8 cell diameters over five divisions (Figs 5D, S10B, grey). This constant cluster size despite the state propagations during the cell divisions is maintained by the short-range communication that induces cell type transitions following each division in the simulation. The cluster radius in simulations is thus broadly consistent with the experimentally measured values at early, but not at late stages of differentiation.

Cluster formation at later stages of differentiation could be driven by long-range communication via FGF4, or by FGF4-independent mechanisms. To distinguish between these possibilities, we analysed the spatial arrangement of cell types in rescued *Fgf4* mutant cells, using 4 h or 8 h of GATA4-mCherry induction together with differentiation in the presence of 10 ng/ml or 2.5 ng/ml FGF4, respectively, to obtain similar cell type proportions as in the wild type (Figs 2B, 4C). In both *Fgf4* mutant conditions, cell types were initially well mixed, and clustered at later stages of differentiation (Figs 5A and S9B–D). Both the median cluster radius as well as its increase between 16 h and 40 h were similar in wild type and in *Fgf4* mutant cells (Figs 5B,C, blue, S11). The continuous increase in cluster size was recapitulated in simulations of the mutant case. In these simulations, cluster sizes were initially comparable to the wild type case, but increased rapidly as cells divided, since in the absence of coupling-dependent cell type transitions gene expression states and thus cell types propagated locally (Figs 5D, S10, blue). Consequently, the transition from a well-mixed to a clustered cell type arrangement observed experimentally in the wild type could be recapitulated if following the settling to committed cell types, the communication was removed from the system (Figs 5D, S10, orange).

Taken together, these results suggest that at early stages, the spatial organization of cell types is consistent with regulation by a local cell-cell communication mechanism. Later on however, additional FGF4-independent mechanisms such as cell division and active cell sorting dominate the spatial organization.

Development • Accepted manuscript

Heterogeneous differentiated cell types are maintained by intercellular communication

A central characteristic of a population-based mechanism for cell differentiation, such as the IHSS, is the interdependence of different cell types (Koseska and Bastiaens, 2017). It has been theoretically demonstrated that this property of the IHSS solution manifests in the regeneration of heterogeneous populations following separation of cell types after differentiation (Stanoev et al., 2021). To test whether a similar behaviour could be observed in isolated PrE-like cells, we used a short-lived Venus-NLS-PEST reporter (Abranches et al., 2013; Nagoshi et al., 2004) knocked into the *Gata6* locus of a GATA4-mCherry inducible cell line as a proxy to isolate PrE-like cells, and to subsequently monitor their differentiation state (Fig. 6A). When putative PrE-like cells were isolated by flow sorting for VNP expression 16 h after the end of a doxycycline pulse and cultured in N2B27 medium, they regenerated a mixture of VNP-positive and –negative cells within 10 h, resembling cell colonies that had not been disrupted and sorted (Fig. 6B–D, first and second row). The smaller proportion of VNP-positive cells detected by flow cytometry in the unperturbed cultures is likely due to insufficient induction of the GATA4-mCherry transgene in this cell line (Fig. 6D). While VNP expression was stable over time in most cells in unperturbed colonies, in sorted VNP-positive cells growing in N2B27, the reporter was first globally downregulated before the heterogeneous expression patterns emerged (Fig. 6C, first and second panel, Movies S1, S2). Similar transitions have been predicted *in silico* as a generic feature of the IHSS solution (Stanoev et al., 2021). The transient downregulation of VNP expression was not observed when supplementing the culture medium with FGF4, but instead reporter expression was maintained in the majority of cells (Fig. 6B–D, third row, Movie S3). Inhibition of FGF/ERK signalling with the MEK inhibitor PD03 in contrast resulted in the rapid downregulation of VNP expression following sorting in all cells (Fig. 6B–D, fourth row, Movie S4). Reporter expression was likewise downregulated in sorted VNP-positive *Fgf4* mutant cells upon culture in N2B27 alone (Fig. S12). These data indicate that cell-cell communication via FGF4/ERK regulates the re-establishment of a mixture of heterogeneous cell types in a population. *Fgf4* mRNA expression dynamics in sorted VNP-positive further supported this idea (Fig. 6E). *Fgf4* transcripts could hardly be detected immediately after sorting, consistent with the repression of *Fgf4* by endogenous GATA6 in VNP-positive cells. 6 h later, when the VNP reporter and hence endogenous GATA6 expression had dropped, some cells started to re-express *Fgf4* transcripts. 10 h after sorting, a subset of cells re-expressed *Gata6* mRNA and VNP, leading to the mutually exclusive expression of VNP, *Fgf4* and *Gata6* mRNA similar to the situation before sorting (Fig. 6E). Taken together, these

Development • Accepted manuscript

results indicate FGF/ERK signalling re-establishes populations with different cell types following the isolation of PrE-like cells. In unperturbed cell colonies, intercellular communication via FGF/ERK therefore not only generates, but also actively maintains balanced proportions of differentiated cell types.

Discussion

Here we report emergent population-level behaviour during the differentiation of Epi- and PrE-like cells from ESCs expressing inducible GATA factors: Robust proportions of the two cell types are specified from a wide range of GATA induction levels, and re-established from isolated PrE-like cells. This collective behaviour relies on local cell-cell communication via FGF4. The observed differentiation characteristics recapitulate the properties of a population-based dynamical solution, an inhomogeneous steady state, recently proposed as a generic mechanism underlying robust differentiation (Stanoev et al., 2021). Our results suggest a new function for FGF signalling, which is to generate and maintain robust proportions of differentiated cell types.

In contrast to previous studies that reported PrE-like differentiation in *Fgf4* mutant ESCs upon permanent high-level expression of exogenous GATA factors (Kang et al., 2012; Wamaitha et al., 2015), we find that PrE-like cells do not differentiate from *Fgf4* mutant ESCs upon transient GATA induction. This recapitulates the *Fgf4* mutant phenotype in the embryo (Feldman et al., 1995; Kang et al., 2012; Krawchuk et al., 2013). Both in *Fgf4* mutant embryos and ESCs, cell type proportions can be controlled by recombinant FGF4 in a dose-dependent manner (Krawchuk et al., 2013; Yamanaka et al., 2010). Furthermore, the differentiation of Epi- and PrE-like cells *in vitro* recapitulates the remarkably constant proportions of cell types seen in the developing embryo (Saiz et al., 2016). Lastly, cell identities in ESC populations are plastic and can be re-specified upon changing a cell's environment, again similar to observations in the embryo (Arias et al., 2013; Grabarek et al., 2012). Although the specification of PrE-like cells in ESCs does not occur spontaneously and requires the use of inducible transgenic GATA factors, the parallels between proportioning of Epi- and PrE-like cells in ESCs and the patterning of the ICM of the mouse preimplantation embryo suggests that similar mechanisms operate in both systems. Consistent with this idea, a recent study using chimaeras and targeted ablation of specific cell types concluded that an FGF4-based population-level mechanism balances the size of the Epi and the PrE lineage in the mouse embryo (Saiz et al., 2020). In ESCs, this population level mechanism manifests in

Development • Accepted manuscript

defined media without extrinsic growth factors, in contrast to previously used culture conditions that supply exogenous signals and thereby reveal the dynamics of isolated cell-intrinsic regulatory circuits (Schröter et al., 2015). The robust proportioning of Epi- and PrE-like cells in defined minimal medium thus provides another example how endogenous signalling interactions lead to robust patterning of ESC populations (Turner et al., 2017).

The ESC system allowed identifying new regulatory links of the communication mechanism, and testing functional properties that were so far inaccessible. First, our data suggests that the direct repression of *Fgf4* by GATA factors communicates cell state to the population through a reduction of FGF4 signalling. In ESCs, this new regulatory link appears to predominate over previously proposed mechanisms such as the indirect regulation of *Fgf4* expression through NANOG, or the regulation of *Fgfr* expression by GATA factors (Frankenberg et al., 2011; Wamaitha et al., 2015). Second, our ability to experimentally control transgenic GATA levels in ESCs allowed us to show that cell-cell communication buffers cell type proportions against a broad range of initial conditions as predicted by theory. Third, chimeric cultures of wild type and *Fgf4* mutant cells revealed that cell-cell communication via FGF4 signals acts over a short spatial range. This finding suggests a mechanistic explanation for the spatially random differentiation of Epi and PrE cells in the ICM (Chazaud et al., 2006; Fischer et al., 2020; Plusa et al., 2008; Rossant et al., 2003). The short activity range of FGF4 requires cells of opposite fates to be closely juxtaposed until cell fates become irreversibly determined, such that the formation of spatially segregated tissues needs to be achieved through a subsequent sorting step. Our analysis of spatial patterns in ESC cultures indicates that this sequence of first deploying communication via FGF4 to establish spatially intermingled robust proportions of cell types followed by an FGF-independent sorting step is conserved *in vitro.*

The repressive coupling of cell fates via a short-range FGF4 during ESC differentiation parallels central features of Delta-Notch signalling during lateral inhibition (Ferrell, 2012; Henrique and Schweisguth, 2019; Hori et al., 2013; Simpson, 1990). Consequently, hallmarks of the population-level behaviour mediated by FGF4 in differentiating ESCs are recapitulated in an engineered cell system where cells communicate via Delta-Notch, such as the differentiation of discrete cell types in reproducible proportions, the re-establishment of those proportions upon removal of one cell type, and the dependence of cell type proportions on cell density or contact (Matsuda et al., 2015). Thus, when connected to appropriate intracellular

Development • Accepted manuscript

regulatory circuits, molecularly diverse intercellular communication systems can yield similar functional outputs.

In rescued *Fgf4* mutant cells that do not communicate, differentiation outcomes in the cell population strongly depend on the distribution of GATA4-mCherry induction levels. This is in line with predictions from single cell models for cell differentiation, in which initial conditions in individual cells strongly influence their differentiation path (Huang et al., 2007). When cells communicate via FGF4 in contrast, the collective differentiation outcome is robust and becomes insensitive to the distribution of GATA4-mCherry induction levels. Thus, the behaviour of the communicating cell population cannot directly be extrapolated from the behaviour of single isolated cells. Theoretically, the conceptual differences between single-cell and population-based modes of differentiation manifest in the emergence of a new type of solution that jointly describes the heterogeneous cell identities, an IHSS, in the communicating cell population. The robust generation of cell type proportions irrespective of initial conditions, and their active maintenance through intercellular communication that we observe experimentally, are two key properties of the IHSS (Stanoev et al., 2021). This suggests that the IHSS is a likely dynamical mechanism underlying the differentiation of cells with discrete identities during mammalian preimplantation development. Given the pervasiveness of robust cell type proportioning during development and homeostasis (Viader-Llargués et al., 2018), it is likely that similar population-based mechanisms underlie canalized development in diverse systems in which multipotent progenitor cells give rise to several differentiated cell types.

Methods

Cell lines

Cell lines used in this study were E14tg2a (Hooper et al., 1987) and an *Fgf4* mutant *Spry4*$^{H2B\text{-}Venus/+}$ line that we have previously described (Morgani et al., 2018). dsRed-labelled cells were from an E14tg2a-background and kindly supplied by J Nichols. The *Gata6*^{VNP} allele reporter was established in the background of a previously described cell line carrying a doxycycline-inducible GATA4-mCherry transgene in the Col1a1 locus as well as a randomly integrated H2B-Cerulean nuclear marker driven by a CAGS promoter (Schröter et al., 2015).

E14tg2a-based inducible cell lines were maintained on fibronectin-coated tissue culture plastic in 2i + LIF medium, which consists of a N2B27 basal medium supplemented with 3

Development • Accepted manuscript

µM CHIR99021 (Tocris), 1µM PD0325901 (SelleckChem) and 10 ng/ml LIF (protein expression facility, MPI Dortmund). For maintenance of *Fgf4* mutant subclones, we supplemented the 2i + LIF medium with 10% fetal bovine serum (FBS), as *Fgf4* mutant lines showed severely decreased proliferation upon long-term culture in 2i + LIF alone. FBS was removed at least one day before the experiment.

Spry4-reporter cell lines to measure signalling range, as well as *Gata6*-reporter cell lines were maintained on gelatin coated dishes in GMEM-based medium supplemented with 10% FBS, sodium pyruvate, 50 µM β-mercaptoethanol, glutamax, non-essential amino acids and 10 ng/ml LIF. 1 µM PD0325901 was added to the cultures of *Spry4*- and *Gata6*-reporters three days before the experiment, to downregulate *Spry4* reporter expression, or to capacitate cells for PrE-like differentiation (Schröter et al., 2015).

FGF4 was from Peprotech and supplied in the indicated concentrations, together with 1 µg/ml heparin (Sigma).

Genetic engineering of ESC lines

Doxycycline-inducible *GATA4*-mCherry inducible ES cells were generated by electroporation of 50.000 E14tg2a ES cells with 4 µg of pPB-TET-GATA4-mCherry, 4 µg pCAG-rtTA-Neo, and 4 µg pCAG-PBase (Wang et al., 2008), followed by G418 selection (400 µg/ml) one day after transfection. We established more than 10 independent clonal lines and assayed induction levels and homogeneity by flow cytometry 2 – 8 h after induction of transgene expression by adding 500 ng/ml doxycycline to the culture medium. Four clones with homogeneous induction levels were chosen and maintained under G418 selection, to circumvent silencing of the inducible transgene.

Mutagenesis of the *Fgf4* locus was performed as previously described (Morgani et al., 2018). *Fgf4* loss of function clones were identified by PCR-amplification, cloning and sequencing of a sequence around the *Fgf4* start codon. We either selected clones with a targeted mutation delivered by a single-stranded DNA repair template that we have previously shown to disrupt *Fgf4* function (Morgani et al., 2018) , or selected at least two independent clones carrying indels around the start codon that introduced frameshift as well as nonsense mutations. All independent clones with random indels showed indistinguishable behaviour in the differentiation assays.

The *Gata6* reporter cell line was generated using previously described knock-out first targeting arms of the EUCOMM project (Skarnes et al., 2011), combined with a VNP reporter cassette (Nagoshi et al., 2004) and a neomycin resistance gene driven from a human β-actin

Development • Accepted manuscript

promoter. This construct was integrated by homologous recombination into a line carrying a doxycycline-inducible GATA4-mCherry transgene in the *Col1a1* locus as well as a randomly integrated H2B-Cerulean nuclear marker driven by a CAGS promoter described in (Schröter et al., 2015). Clones were screened for correct integration of the reporter construct by long range PCR spanning the targeting arms.

The targeting construct to generate the *Spry4*$^{H2B\text{-}Venus}$ allele in GATA4-mCherry inducible cell lines was based on the one used in (Morgani et al., 2018), except that the puromycin selectable marker was exchanged for a neomycin cassette. The construct was integrated into ESCs by homologous recombination. To increase targeting efficiency, cells were co-transfected with a plasmid expressing Cas9 and a sgRNA that targets a sequence at the 5' end of the 5' targeting arm which is present in the endogenous *Spry4* locus, but not the targeting construct. Neomycin-resistant clones were expanded and screened for correct integration of the reporter construct by long range PCR spanning the targeting arms.

Deletion of the putative GATA-binding element upstream of Fgf4 was performed by co-transfecting two plasmids expressing Cas9 and sgRNAs flanking the binding element (FGF4_GATAbind_guide5'-2: 5'-AGGGTCTCTGTTCAGGGACA-3'; FGF4_GATAbind_guide3'-1: 5'-CCACATAAGTACCATAGTAT-3'). Following selection for successful transfection, clonal lines were established and tested via PCR with primers Fgf4_GATAbind_fwd2: 5'-GACAGCAACAGTGGATTCAC-3' and Fgf4_GATAbind_rev2: 5'-ACCCCAGTCTTCTGCAAGAG-3' for the presence of a deletion of expected size. Deletion of the binding site was further confirmed by Sanger sequencing of PCR amplicons.

All genetically modified lines were karyotyped using standard procedures (Nagy et al., 2008), and all except one clonal line (C5) were confirmed to have a median chromosome count of n = 40.

Immunostaining and image analysis

Immunostaining of adherent cells was performed as previously described (Schröter et al., 2015). Antibodies used were anti-NANOG (e-bioscience, eBioMLC-51, 14-5761-80, final concentration 2.5 µg/ml), anti-GATA6 (rabbit polyclonal, Invitrogen, PA1-104, final concentration 5 µg/ml), anti-GATA6 (goat polyclonal, R&D AF1700, final concentration 1 µg/ml), anti-laminin (Sigma-Aldrich L9393, final concentration 0.5 µg/ml), anti-SOX17 (R&D AF1924, final concentration 1 µg/ml), and anti-FLAG (Sigma-Aldrich F1804-200, final concentration 1 µg/ml). Secondary antibodies were from Invitrogen/LifeTech. Images were acquired using a 63x 1.4 N.A. oil-immersion objective on a confocal Leica SP8

Development • Accepted manuscript

microscope, with all settings held constant between replicates. Images were quantified using custom scripts written for ImageJ (NIH) and in MATLAB (The MathWorks).

Flow cytometry

Staining for flow cytometric analysis of intracellular antigens was performed as previously described (Schröter et al., 2015). Primary and secondary antibodies were the same as used for immunostaining. mCherry fluorescence measurements and cell sorting were performed on a BD FACS Aria. All other flow cytometric analysis was carried out using a BD LSR II. Single cell events were gated based on forward and side scatter properties. GATA4-mCherry expression measurements were normalized to the respective uninduced control. For Fig. S1, cell types were assigned using a gaussian mixture model (see below). Otherwise, gates to separate marker-positive from marker-negative cells were determined visually as the threshold that best bisected the bimodal distribution of marker expression across all samples within one experiment.

Membrane labelling

Cell membranes were labelled with CellBrite Fix (Biotium) according to manufacturer's recommendations. Briefly, cells were washed with PBS containing Ca and Mg, and cells were incubated with dye diluted 1:1000 in PBS for 15 minutes at 37ºC. After labelling, cells were washed twice with PBS, followed by fixation and in situ HCR.

In situ HCR and image analysis

Probe sets for *Gata6* and *Fgf4* and corresponding Alexafluor-labelled amplifiers for staining of mRNA molecules via third generation in situ HCR (Choi et al., 2018) were sourced from Molecular Instruments. Staining was performed according to manufacturer's instructions. Briefly, adherent cells were fixed for 15 minutes with 4% paraformaldehyde, washed with PBS and permeabilized for several hours in 70% ethanol at -20ºC. Cells were then washed twice with 2x SSC and equilibrated in probe hybridization buffer for at least 30 minutes. Transcript-specific probes were used at a concentration of 4 nM and hybridized overnight. Excess probe was removed through several washes with probe wash buffer and 5x SSCT, and cells were equilibrated in amplification buffer for at least 30 minutes. Fluorescently labelled amplifiers were used at a concentration of 60 nM. Amplification was allowed to proceed for 16 – 24 h at room temperature. Excess amplifier was removed by several washes with 5x SSCT, followed by counterstaining with Hoechst 33342 and mounting in glycerol-based

Development • Accepted manuscript

medium. Imaging was performed on an SP8 confocal microscope with a 63x (NA1.4) lens. For image analysis, custom scripts written for ImageJ (NIH) and MATLAB (The MathWorks) were used to segment nuclei based on the Hoechst 33342 image, and to identify cells based on the CellBrite membrane stain where available. For images with a cell membrane stain, *Fgf4* mRNA staining intensities were integrated across the entire cell. For cells differentiated for 40 h, where identification of the cell outlines with CellBrite proved difficult, we assigned *Fgf4* mRNA signals to individual cells by dilating nuclear masks. Briefly, nuclei were first identified based on the Hoechst 33342 staining, and then dilated using a non-merge dilation. We then measured total mRNA staining within these dilated nuclear masks.

Analysis of ChIP-seq data

Raw data of GATA6-ChIPseq from Wamaitha et al., 2015, was downloaded from NCBI and mapped to the mouse genome (mm10/GRC38) with BowTie2 and default parameters on galaxy.org. Mappings were visualized with IGV.

Decay length measurements

Fgf4 mutant Spry4$^{H2B\text{-}Venus}$ reporter cells (Morgani et al., 2018) were seeded at a density of $5*10^4$ cells/cm^2 in N2B27. 2 h later, dsRed-expressing cells were added at a density of 500 cells/cm^2. For the first 3 hours of co-culture, the medium was supplemented with 250 to 500 nM siR-Hoechst (Lukinavičius et al., 2015) to label nuclei. 12 h later, live cells were imaged on a Leica SP8 confocal system. Nuclei were segmented in FIJI (Schindelin et al., 2012), and for each Spry4$^{H2B\text{-}Venus}$ reporter cell in the vicinity of a ds-Red expressing cell, the background-subtracted Venus fluorescence intensity as well as the distance to the centre of mass of the dsRed expressing cells was determined. Cells were grouped according to their distance from dsRed expressing cells in 3 µm bins, and mean fluorescence intensities for each bin plotted versus their distance. Decay length was estimated in GraphPad Prism by fitting a plateau followed by a one-phase decay function.

Cell differentiation in chimeric cultures

Fgf4 mutant cells carrying an inducible GATA4-mCherry transgene were seeded at a density of $5*10^4$ cells/cm^2. 24 h later, cells were induced for 8 h with 500 ng/ml dox in 2i + LIF medium. Medium was switched to N2B27 concomitantly with the addition of 5000 wild type cells per cm^2 expressing an ERK-KTR-Clover protein (Raina et al., 2020) as cell label.

Development • Accepted manuscript

Cultures were fixed 16 h later, stained for GATA6, NANOG, GFP and DNA, and both colonies with and without labelled wild type cells from the same well were imaged on a Leica Sp8 confocal microscope. For analysis, we used the blind analysis tool in FIJI. The number of GATA6+;NANOG- cells was scored for each colony without knowing whether the colony contained *Fgf4* wild type, and results were grouped according to the presence or absence of labelled *Fgf4* cells after completion of scoring.

Determination of cell-cell distances

Neighbourhood graphs were constructed for each field of view by using the cell positions to generate a Delaunay triangulation and the corresponding Voronoi diagram. Spurious links between non-adjacent cells were trimmed by excluding links in the Delaunay graph that do not directly pass through the shared Voronoi edge between the two respective Voronoi cells. Links between adjacent cells (purple links in Fig S6A) were pooled to generate a distance distribution between nearest neighbours (purple histogram, Fig S6B). The distances of links between unconnected cells that share a neighbour (yellow in Fig S6A), were likewise pooled to give the distribution of second-nearest neighbour distances (yellow histogram, Fig S6B). Both distributions were independently fit with a 2-component Gaussian mixture model, to separate the true distributions of nearest and second-nearest neighbour distances from higher-order distributions arising from erroneously assigned links (purple and yellow dashed lines, Fig S6B). The mean and SD of nearest- and second-nearest neighbour distances was estimated from the first component of these Gaussian fits.

Assignment of cell types with Gaussian mixture model

To characterize the emergence of the two main cell types (Fig. S1) and for spatial clustering analysis, we focused on GATA6+; NANOG- (G+) and GATA6-; NANOG+ (N+) cells only. To identify these cell types specifically, we applied a two-component Gaussian mixture model (GMM) fit to single cell distributions in the GATA6; NANOG expression space, using the MATLAB function fitgmdist(). The fitted GMM assigns each cell a posterior probability of its association with one of the two component distributions. Cells with a posterior probability of >0.9 for one of the two components were classified to the corresponding cell type. When applying this approach to the flow cytometry data shown in Fig. S1, we also excluded outlier cells. To define outliers for each population of cell types, we first measured the Euclidean distance of each point from its respective population centre, defined as the

Development • Accepted manuscript

mean of the GMM fit. Next, we eliminated cells with a Euclidean distance of greater than 4 inter-quartile ranges from the respective population centre.

Analysis of neighbourhood composition

To determine local neighbourhood composition, we first generated a matrix of Euclidean distances between every cell classified as either G+ or N+ in each field of view. For every cell, we then computed the fraction of G+ and N+ cells within a Euclidean distance of 31.7 µm, encompassing most of the nearest and second nearest neighbours. Random distributions (blue lines in Fig. S9) were calculated by populating the local neighbourhood of each cell randomly from the global distribution in each treatment.

Measurement of cluster radius

The average cluster radius of N+ cells was estimated with the same method in experimental data where cell types had been categorized with a Gaussian mixture model, and in simulations of the model. We calculated for all N+ cells in a field of view or simulation run the decrease in the fraction of N+ cells in neighbourhoods of increasing radius, settling to the overall fraction of N+ cells. For scaling, we first subtracted the overall fraction of N+ cells in the respective field of view or simulation run, and then normalized values such that the scaled fraction at zero distance was set to one again. The cluster radius is the distance at which the scaled fraction of N+ cells is equal to 0.5.

Statistical analysis

Significance testing for cell type proportions was performed in GraphPad Prism, using one-way ANOVA for matched data with Gaussian distribution, followed by Tukey's or Dunnett's multiple comparison test or a test for a linear trend. In all other conditions, student's t-test was used in MATLAB.

Live cell imaging and tracking

To track *Gata6* reporter expression in live cells, PrE-like differentiation was induced by a 6 h pulse of doxycycline-treatment in serum-containing medium as described in (Schröter et al., 2015). 16 hours after doxycycline-removal, cells were either switched directly to N2B27 medium lacking phenol red, or trypsinised, sorted for reporter expression, and seeded on fibronectin-coated imaging dishes (ibidi µ-slides). Time-lapse imaging was started within 2 h after sorting on an Olympus IX81 widefield microscope equipped with LED illumination

Development • Accepted manuscript

(pE4000, CoolLED) and a Hamamatsu c9100-13 EMCCD camera. Hardware was controlled by MicroManager software (Edelstein et al., 2001). Time-lapse movies were acquired using a 40x oil immersion lens (NA 1.2), with 10-minute time intervals.

Cell tracking was carried out with TrackMate (Tinevez et al., 2017) based on the constitutively expressed H2B-Cerulean nuclear marker. Fluorescence intensity was measured in a circular region of interest in the centre of the nucleus, and background-subtracted fluorescence intensities plotted in Python. Trace colour in Fig. 6D was assigned according to fluorescence intensity in the last frame of the movie, with respect to the estimated intensity threshold used for flow sorting (dashed line).

Computational model for cell type proportioning

The model of the intercellular communication system (Fig. 4D) is adapted from Stanoev et al. (2021), and is described with the following set of equations:

$$\frac{1}{\lambda}\frac{dN_i}{dt} = \alpha_N \frac{1}{1+G_i^{\beta}} + \alpha_{N,F}\frac{1}{1+F_{ext,i}^{\eta}} - N_i$$

$$\frac{1}{\lambda}\frac{dG_i}{dt} = \alpha_G \frac{1}{1+N_i^{\gamma}} - G_i$$

$$\frac{1}{\lambda}\frac{dF_i}{dt} = \alpha_F \frac{1}{1+G_i^{\delta}} - F_i$$

N_i and G_i describe NANOG and GATA6 protein expression levels in cell i, regulated by mutual inhibition, while F_i is the secreted FGF4 whose production is downregulated by GATA6. $F_{ext,i} = \frac{1}{|N(i)|+1}\sum_{j\in(N(i)\cup i)} F_j$ is the extracellular FGF4 concentration that is sensed by cell i from its neighbourhood $N(i)$, resulting in downregulation of NANOG production in the cell. $\alpha_N = 2.5, \alpha_{N,F} = 0.5, \alpha_G = 3$ and $\alpha_F = 3$ denote production rate constants, $\beta = \eta = \gamma = \delta = 2$ are the Hill coefficients, degradation rates were set to 1 as $\lambda = 50$ was used as a scaling kinetic parameter. 10000 cells were deployed on a regular 100x100 two-dimensional lattice with no-flux boundary conditions. Cell-cell communication was modelled to be short-range, reflecting the experimental wild-type case, i.e. communication between direct neighbours and cells two hops away on the lattice. When mimicking the *Fgf4* mutant case, communication between cells was excluded, and an external input was modelled with $F_{ext} = 1.2$ for the results shown in Fig. 4E-G. For Fig. S7B, F_{ext} was varied as indicated. When supplementing external input in the wildtype case, $F_{ext} = 1.0$ was added to communication input $F_{ext,i}$ for each cell i.

Development • Accepted manuscript

The cell populations were initiated analogously to the experimental case, by varying the initial conditions of all cells from being NANOG-expressing, through intermediate NANOG and GATA6 expression, to being GATA6-expressing (Fig. 4E, left column). More specifically, the variables were sampled independently from unimodal Gaussian distributions $\mathcal{N}(\mu_{ics}(p), \sigma_{ics} = 0.1 * \mu_{ics}(p))$, with the mean $\mu_{ics}(p) = (1-p) * \mu_{G-;N+} + p * \mu_{G+;N-}$ placed on the line segment connecting the GATA6-; NANOG+ state $\mu_{G-;N+}$ and the GATA6+; NANOG- state $\mu_{G+;N-}$, partitioning it in proportion p. $p \in \{0, 0.4, 0.5, 0.6, 1\}$ was used for the quantifications in Fig. 4G. Samples from around the endpoints and the midpoint ($p \in \{0, 0.5, 1\} \Rightarrow \mu_{ics} \in \{\mu_{G-;N+}, \frac{1}{2}(\mu_{G-;N+} + \mu_{G+;N-}), \mu_{G+;N-}\}$) are shown in Fig. 4E, left column.

Cell heterogeneity was introduced by varying all of the parameters independently with standard deviation of 0.02 from the respective values for each cell. Stochastic differential equation model was constructed from the deterministic equations by adding a multiplicative noise term $\sigma X dW_t$, where dW_t is the Brownian motion term, X is the variable state and $\sigma = 0.1$ is the noise term. The model was solved with $\Delta t = 0.01$ using the Milshtein method (Mil'shtejn, 1975). Following integration, cell identities were estimated by comparing the NANOG and GATA6 values from the final states of the cells, and the ratios were computed.

For comparing the spatial organizations between communicating and non-communicating cells at different time points, periodic synchronous cell divisions in the population were included in the model as in (Stanoev et al., 2021), spanning 5 cell cycles. Cell divisions occur along the horizontal and vertical axes on the grid alternately, sequentially yielding lattices of 10x10, 10x20, 20x20, 20x40 and 40x40 (Fig. S10A). At every cell division, the final state of the mother cell is passed on to the daughter cells' initial conditions. The mother cell's parameter set is also inherited. Spatial organizations were analysed at the end of each cell cycle, after the collective state is allowed to reach a steady state in a deterministic fashion, by estimating the N+ cluster radius as described above.

For the hybrid model, it was assumed that after the third cell cycle the cells commit to their current fates and the communication becomes inconsequential, effectively bringing about a switch to a non-communicating grid. For all conditions, cells' states were initialized with Gaussians with $\mu_{ics}(0.5)$, as described above.

The numerical bifurcation analysis for the 2-cell system (Fig. S8) was performed using the XPP/AUTO software (http://www.math.pitt.edu/~bard/xpp/xpp.html). All simulations were performed using custom-made code in MATLAB (MATLAB and Statistics Toolbox Release R2020b, The MathWorks).

Development • Accepted manuscript

Acknowledgements

We thank J. Nichols for sharing dsRed-labelled ESCs, L. Süther for help with generating inducible cell lines, M. Schulz and S. Müller for assistance with microscopy and flow cytometry, and P. Bastiaens for stimulating discussion and conceptual input on the project. We thank P. Bieling, J. de Navascues, G. Vader, S. Fischer and the members of the Schröter and Koseska groups for discussions and comments on the manuscript. All authors are supported by the Max Planck Society.

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Conceptualization, A.K. and C.S.; Methodology, A.S. and C.S.; Investigation, D.R., A.B., A.S., M.P., and C.S.; Validation, D.R., A.B., and M.P.; Visualization, D.R., A.B., and A.S.; Formal analysis, D.R., A.B., and A.S.; Writing – original draft, C.S.; Writing – review and editing, all authors; Supervision, A.K. and C.S..

References

Abranches, E., Bekman, E. and Henrique, D. (2013). Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. *PLoS ONE* 8, e59928.

Arias, A. M., Nichols, J. and Schröter, C. (2013). A molecular basis for developmental plasticity in early mammalian embryos. *Development (Cambridge, England)* 140, 3499–3510.

Bedzhov, I., Graham, S. J. L., Leung, C. Y. and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philosophical Transactions Royal Soc B Biological Sci* 369, 20130538.

Bessonnard, S., Mot, L. D., 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 (Cambridge, England)* 141, 3637–3648.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255–256.

Development • Accepted manuscript

Caluwé, J. D., Tosenberger, A., Gonze, D. and Dupont, G. (2019). Signalling-modulated gene regulatory networks in early mammalian development. *Journal of theoretical biology* 463, 56–66.

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. *Developmental Cell* 10, 615–624.

Chickarmane, V. and Peterson, C. (2008). A Computational Model for Understanding Stem Cell, Trophectoderm and Endoderm Lineage Determination. *PLoS ONE* 3, e3478-8.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A. and Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753.

Collier, J. R., Monk, N. A., Maini, P. K. and Lewis, J. H. (1996). Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *Journal of theoretical biology* 183, 429–446.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R. and Stuurman, N. (2001). *Computer Control of Microscopes Using μManager*. John Wiley & Sons, Inc.

Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science (New York, NY)* 267, 246–249.

Ferrell, J. E. (2012). Bistability, Bifurcations, and Waddington's Epigenetic Landscape. *Curr Biol* 22, R458–R466.

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.

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. *Developmental Cell* 21, 1005–1013.

Gardner, R. L. (1968). Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature* 220, 596–597.

Grabarek, J. B., Zyzynska, K., Saiz, N., Piliszek, A., Frankenberg, S., Nichols, J., Hadjantonakis, A.-K. and Plusa, B. (2012). Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. *Development (Cambridge, England)* 139, 129–139.

Hamilton, W. B. and Brickman, J. M. (2014). Erk Signaling Suppresses Embryonic Stem Cell Self-Renewal to Specify Endoderm. *Cell Reports* 9, 2056–2070.

Henrique, D. and Schweisguth, F. (2019). Mechanisms of Notch signaling: a simple logic deployed in time and space. *Development (Cambridge, England)* 146, dev172148-11.

Development • Accepted manuscript

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* 326, 292–295.

Hori, K., Sen, A. and Artavanis-Tsakonas, S. (2013). Notch signaling at a glance. *Journal of Cell Science* 126, 2135–2140.

Huang, S., Guo, Y.-P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology* 305, 695–713.

Kang, M., Piliszek, A., Artus, J. and Hadjantonakis, A. K. (2012). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development (Cambridge, England)* 140, 267–279.

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.

Koseska, A. and Bastiaens, P. I. (2017). Cell signaling as a cognitive process. *The EMBO journal* 36, 568–582.

Krawchuk, D., Honma-Yamanaka, N., Anani, S. and Yamanaka, Y. (2013). FGF4 is a limiting factor controlling the proportions of primitive endoderm and epiblast in the ICM of the mouse blastocyst. *Developmental Biology* 384, 65–71.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* 134, 2895–2902.

Lukinavičius, G., Blaukopf, C., Pershagen, E., Schena, A., Reymond, L., Derivery, E., González-Gaitán, M., D'Este, E., Hell, S. W., Gerlich, D. W., et al. (2015). SiR-Hoechst is a far-red DNA stain for live-cell nanoscopy. *Nature Communications* 6, 8497.

Matsuda, M., Koga, M., Woltjen, K., Nishida, E. and Ebisuya, M. (2015). Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat Commun* 6, 6195.

Mil'shtejn, G. N. (1975). Approximate Integration of Stochastic Differential Equations. *Theory of Probability & Its Applications* 19, 557–562.

Molotkov, A., Mazot, P., Brewer, J. R., Cinalli, R. M. and Soriano, P. (2017). Distinct Requirements for FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency. *Dev Cell* 41, 511-526.e4.

Morgani, S. M., Saiz, N., Garg, V., Raina, D., Simon, C. S., Kang, M., Arias, A. M., Nichols, J. N., Schröter, C. and Hadjantonakis, A.-K. (2018). A Sprouty4 reporter to monitor FGF/ERK signaling activity in ESCs and mice. *Developmental Biology* 441, 104–126.

Mot, L. D., Gonze, D., Bessonnard, S., Chazaud, C., Goldbeter, A. and Dupont, G. (2016). Cell Fate Specification Based on Tristability in the Inner Cell Mass of Mouse Blastocysts. *Biophysj* 110, 710–722.

Development • Accepted manuscript

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F. and Schibler, U. (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* 119, 693–705.

Nagy, A., Gertsenstein, M., Vintersten, K. and Behringer, R. (2008). Karyotyping Mouse Cells. *Cold Spring Harb Protoc* 2008, pdb.prot4706.

Niakan, K. K., Ji, H., Maehr, R., Vokes, S. A., Rodolfa, K. T., Sherwood, R. I., Yamaki, M., Dimos, J. T., Chen, A. E., Melton, D. A., et al. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes & Development* 24, 312–326.

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 (Cambridge, England)* 135, 3081–3091.

Raina, D., Fabris, F., Morelli, L. G. and Schröter, C. (2020). Intermittent ERK oscillations downstream of FGF in mouse embryonic stem cells. *Biorxiv* 2020.12.14.422687.

Rossant, J., Chazaud, C. and Yamanaka, Y. (2003). Lineage allocation and asymmetries in the early mouse embryo. *Philosophical Transactions Royal Soc Lond Ser B Biological Sci* 358, 1341–1349.

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. *Nature Communications* 7, 13463.

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, 6289.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods* 9, 676–682.

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

Simon, C. S., Hadjantonakis, A.-K. and Schröter, C. (2018). Making lineage decisions with biological noise: Lessons from the early mouse embryo. *Wiley Interdisciplinary Reviews: Developmental Biology* 61, e319-16.

Simpson, P. (1990). Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of Drosophila. *Development (Cambridge, England)* 109, 509–519.

Skarnes, W. C., Rosen, B., West, A. P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A. O., Thomas, M., Harrow, J., Cox, T., et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474, 337–342.

Development • Accepted manuscript

Stanoev, A., Schröter, C. and Koseska, A. (2021). Robustness and timing of cellular differentiation through population-based symmetry breaking. *Development* 148, dev197608.

Tarkowski, A. K. (1959). Experiments on the development of isolated blastomers of mouse eggs. *Nature* 184, 1286–1287.

Tarkowski, A. K. (1961). Mouse chimaeras developed from fused eggs. *Nature* 190, 857–860.

Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., Bednarek, S. Y., Shorte, S. L. and Eliceiri, K. W. (2017). TrackMate: An open and extensible platform for single-particle tracking. *Methods (San Diego, Calif)* 115, 80–90.

Turner, D. A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C. R., Hayward, P. C., Collignon, J., Gustavsen, C., Serup, P., et al. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. *Development* 144, 3894–3906.

Viader-Llargués, O., Lupperger, V., Pola-Morell, L., Marr, C. and López-Schier, H. (2018). Live cell-lineage tracing and machine learning reveal patterns of organ regeneration. *eLife* 7, e08201.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.

Wamaitha, S. E., Valle, I. del, Cho, L. T. Y., Wei, Y., Fogarty, N. M. E., Blakeley, P., Sherwood, R. I., Ji, H. and Niakan, K. K. (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes & Development* 29, 1239–1255.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A. and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences* 105, 9290–9295.

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

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523.

Development • Accepted manuscript

Figures

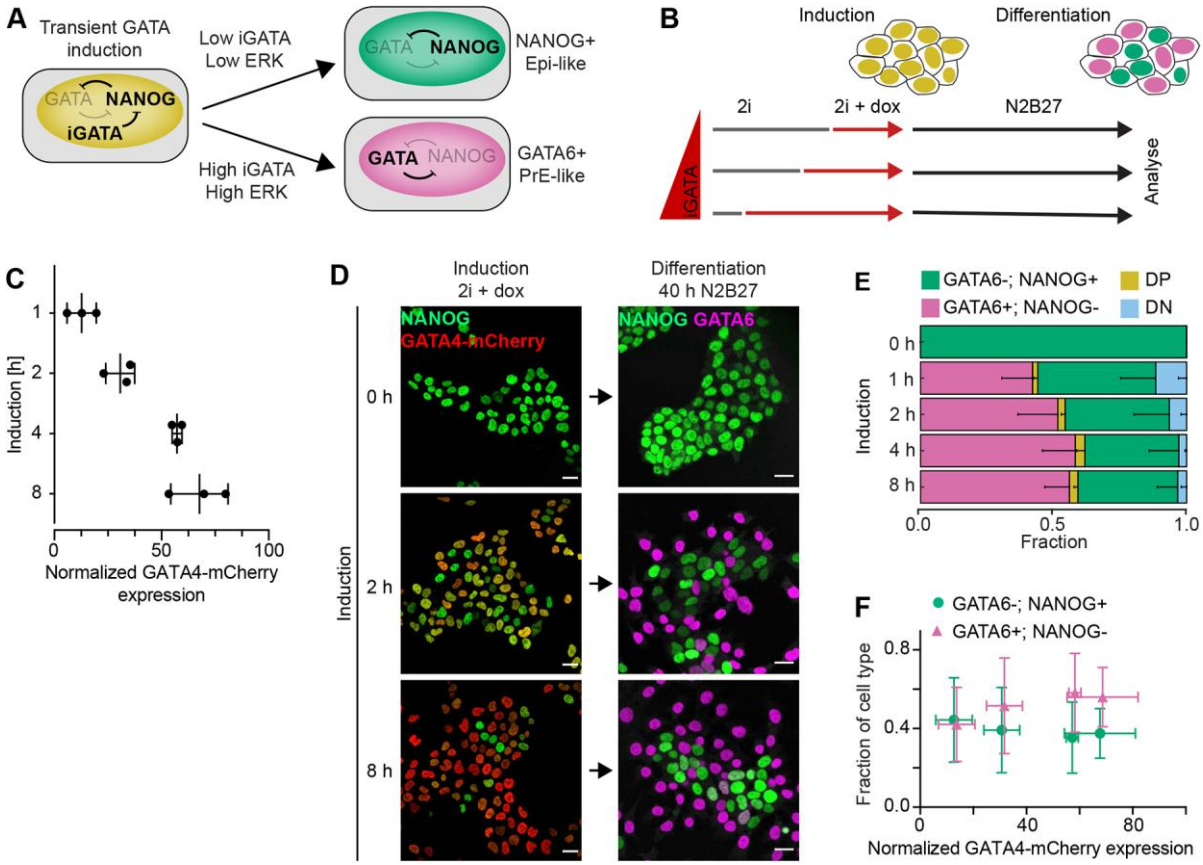


Fig. 1. Proportions of differentiated cell types are independent from GATA4-mCherry induction levels

A Schematic representation of cell differentiation following transient doxycycline-controlled expression of inducible GATA factors (iGATA). **B** Experimental protocol for titrating inducible GATA4-mCherry expression levels by doxycycline addition to individual samples at different time points. The total time from seeding to analysis is held constant. **C** GATA4-mCherry expression levels for different durations of doxycycline induction in 2i + LIF medium measured by flow cytometry, normalized to the non-induced control. Individual data points show mean fluorescence intensities from at least 20000 cells in an individual experiment, bars indicate mean ± SD across N=3 independent experiments. **D** Left: immunostaining for NANOG (green) and GATA4-mCherry (red) in inducible cell lines immediately after the end of a doxycycline pulse of the indicated durations. Right: Immunostaining for NANOG (green) and GATA6 (magenta) in cells treated with doxycycline for the indicated durations, followed by 40 h of differentiation in N2B27 medium. Cells without doxycycline induction have been continuously maintained in 2i medium. **E** Average cell type proportions from N = 4 independent experiments; fraction of GATA6+, NANOG- cells in magenta, GATA6-, NANOG+ cells in green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. Error bars: 95% confidence interval (CI). **F** Plot of average proportions of GATA6+, NANOG- cells (magenta) and GATA6-, NANOG+ cells (green) vs. mean GATA4-mCherry levels for different doxycycline induction times.

Development • Accepted manuscript

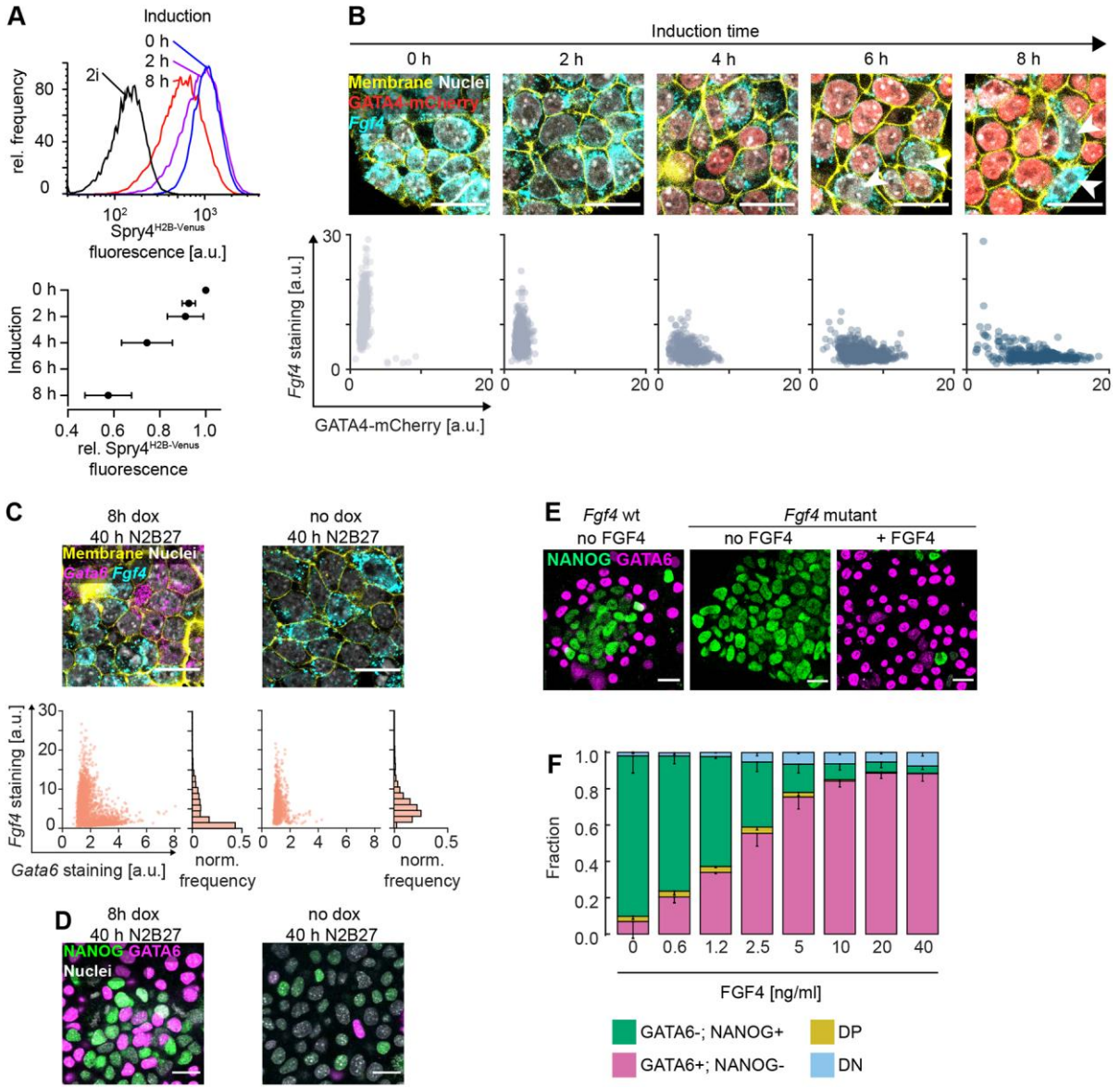


Fig. 2. Differentiating ESCs communicate via FGF4

A Top: Flow cytometry histograms showing Spry4$^{H2B\text{-}Venus}$ reporter expression after 24 h of differentiation in N2B27 medium following the indicated durations of doxycycline induction. Black line indicates reporter expression in cells maintained in 2i medium. Bottom: Mean ± SD of reporter expression from N = 4 independent experiments, normalized to fluorescence levels of cells transferred to N2B27 without doxycycline induction. **B** Top: GATA4-mCherry protein (red) and Fgf4 mRNA expression (cyan) in inducible cells at indicated durations of doxycycline induction. Bottom: Corresponding single cell quantifications. **C** Top: Gata6 (magenta) and Fgf4 mRNA (cyan) expression in inducible cells after 40 h of culture in N2B27 following an 8 h doxycycline pulse (left) or following transfer to N2B27 without induction (right). Bottom: Corresponding single-cell quantifications. Cell membranes and nuclei in **B**, **C** labelled with CellBrite (yellow) and Hoechst33342 (white), respectively. **D** Immunostaining of cells treated as in **C** for GATA6 (magenta) and NANOG (green). Nuclei stained with Hoechst33342 (white). **E** Immunostaining for GATA6 (magenta) and NANOG

Development • Accepted manuscript

(green) in wild type (left) and Fgf4 mutant cells differentiated for 40 h in N2B27 without (middle) or with (right) 10 ng/ml FGF4 after an 8 h doxycycline pulse. Scale bars in **B - E**, 20 μm. **F** Average proportions of cell types in Fgf4 mutant cells induced with doxycycline for 8 h, followed by differentiation in N2B27 in the presence of the indicated concentrations of FGF4. GATA6 and NANOG expression were detected by immunostaining and measured by flow cytometry (see Fig. S5). N = 4; fraction of GATA6+; NANOG- cells in magenta, GATA6-; NANOG+ cells - green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. Error bars: 95% CI.

Development • Accepted manuscript

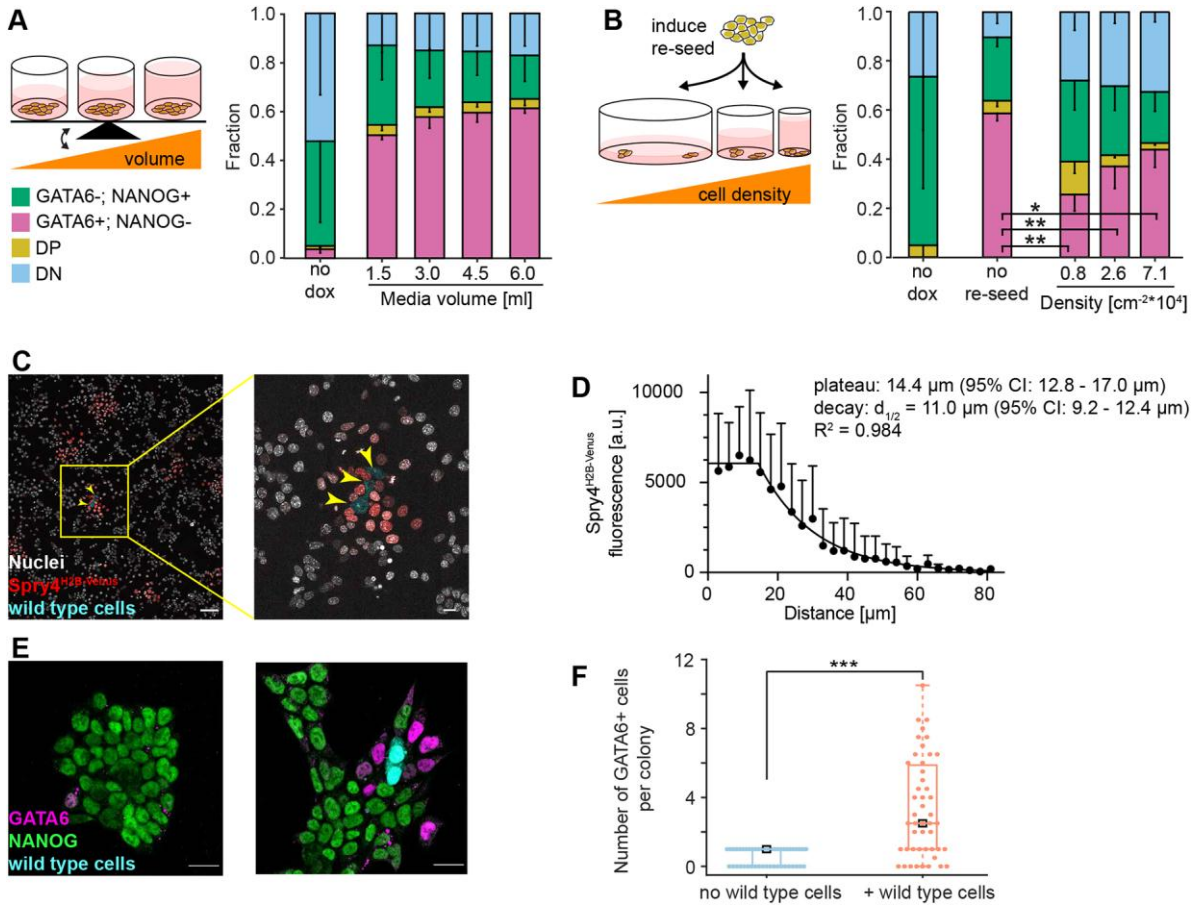


Fig. 3. Communication via FGF4 is spatially restricted

A Left: Experimental approach to test effects of media volume on cell type proportions. Right: Average cell type proportions determined by flow cytometry, N = 3. **B** Left: Experimental approach to test effects of cell density on cell type proportions. Right: Average cell type proportions determined by flow cytometry, N = 4. ** and * indicate $p < 0.005$ and 0.05, respectively (one-way ANOVA). **A**, **B** Fraction of GATA6+; NANOG- cells in magenta, GATA6-; NANOG+ cells in green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. Error bars: 95% CI. **C** Single labelled wild type cells (cyan, yellow arrowheads) seeded on a layer of Fgf4 mutant Spry4$^{H2B\text{-}Venus}$ transcriptional reporter cells. Nuclei are labelled by siR-Hoechst (white), H2B-Venus fluorescence in red. Scale bars, 100 µm (left) and 20 µm (right). **D** Quantitative analysis of FGF4 signalling range. Data points show mean ± SD of background-subtracted H2B-Venus fluorescence intensities in nuclei of Spry4$^{H2B\text{-}Venus}$ reporter cells per distance bin. Data from n = 9 independent signalling centres. The fluorescence decay length was estimated by fitting a plateau followed by one-phase exponential decay to the data (black line). **E** Immunostaining for GATA6 (magenta) and NANOG (green) in chimeric cultures 16 h after an 8 h doxycycline pulse and addition of wild type cells. Staining for GFP (cyan) distinguishes wild type from Fgf4 mutant inducible cells. One representative image of colonies without (left) or with wild type cells (right) is shown. Scale bar, 20 µm. **F** Quantification of GATA6-positive cells in colonies without or with wild type cells ($N = 2$, $n \geq 21$ for each colony type, *** indicates $p < 0.005$, student's t-test, two-tailed, unequal variance).

Development • Accepted manuscript

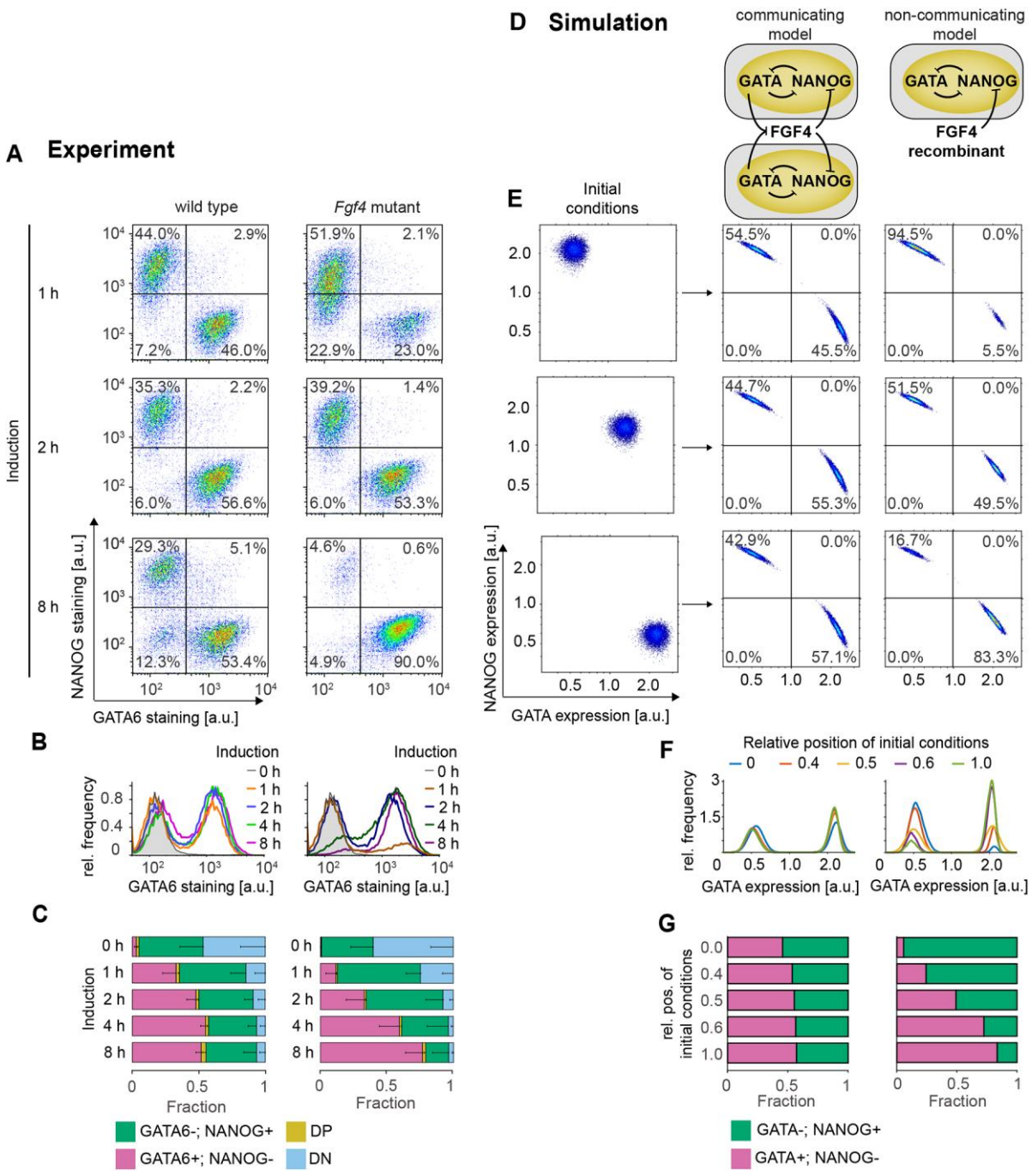


Fig. 4. Cell-cell communication via FGF4 mediates cell type proportioning

A–C Cell differentiation in wild type (left) and Fgf4 mutant cells (right) after doxycycline pulses of indicated duration, followed by 40 h of differentiation in N2B27 alone (wild type) or in N2B27 supplemented with 10 ng/ml FGF4 (Fgf4 mutant). **A** Flow cytometry profiles of NANOG and GATA6 expression. Lines indicate gates to assign cell types. **B** Marginal distributions of GATA6 staining. **C** Quantification of average cell type proportions. GATA6+; NANOG- cells – magenta, GATA6-; NANOG+ cells - green, double positive cells (DP) - yellow, and double negative cells (DN) - blue. N = 4, error bars: 95% CI. **D** Schematic representation of the model with (left) and without cell-cell communication (right). **E** - **G** Influence of initial conditions (left column in **E**) on (GATA6+;

Development • Accepted manuscript

NANOG-) and (GATA6-; NANOG+) cell type proportions in the model with (middle column) or without cell-cell communication (right column). **E** Distributions of initial conditions (left), and corresponding results of numerical simulations with (middle) or without cell-cell communication (right). Lines in middle and left columns indicate gates to assign cell types. **F** Marginal distributions of GATA expression, equivalent to **B**. **G** Quantification of the cell type proportions obtained from the numerical simulations. See Methods for model details and parameters.

Development • Accepted manuscript

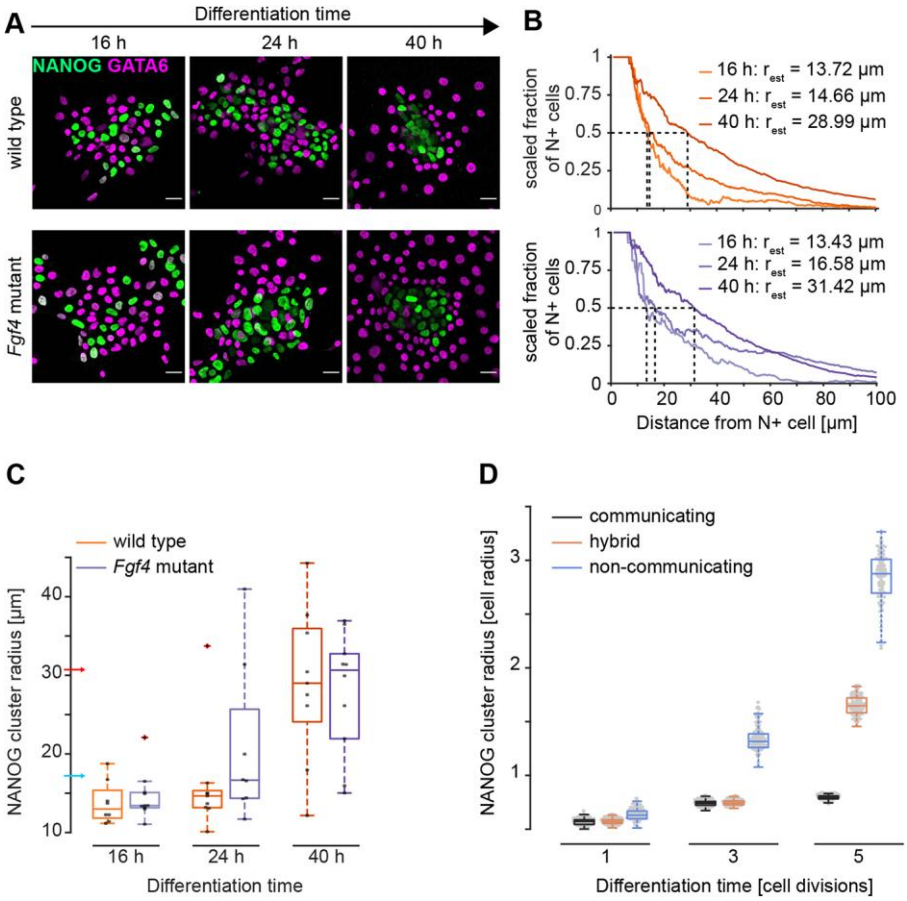


Fig. 5. Spatial arrangement of cell types in wild type and Fgf4 mutant cells.

A Representative immunostainings of wild type and Fgf4 mutant cells for NANOG (green) and GATA6 (magenta) at different time points after the initiation of differentiation. Wild type cells were induced with doxycycline for 8 h and differentiated in N2B27, Fgf4 mutant cells were induced for 4 h and differentiated in N2B27 supplemented with 10 ng/ml FGF4. **B** Estimation of NANOG cluster radius in a single field of view for wild type (orange, top) and Fgf4 mutant cells (blue, bottom) differentiated as in **A**. Graphs depict the scaled fraction of NANOG+ cells within a specific radius around seed cells. Dashed lines indicate determination of cluster radius. **C** Summary statistics of cluster radii for wild type (orange) and Fgf4 mutant cells (blue) differentiated as in **A**. Dots indicate values from individual fields of view, box plots show median, interquartile ranges and outliers (red cross). Blue and red arrowheads indicate mean distance between nearest and second nearest neighbours, respectively (Fig. S5). N = 2, n ≥ 8 for wild type, and N = 1, n ≥ 8 for Fgf4 mutant cells. **D** NANOG+ cluster radii quantified from 100 independent numerical realizations of the model with continuous communication (grey), without communication (blue), and of a hybrid model where communication is switched off after the third division (orange).

Development • Accepted manuscript

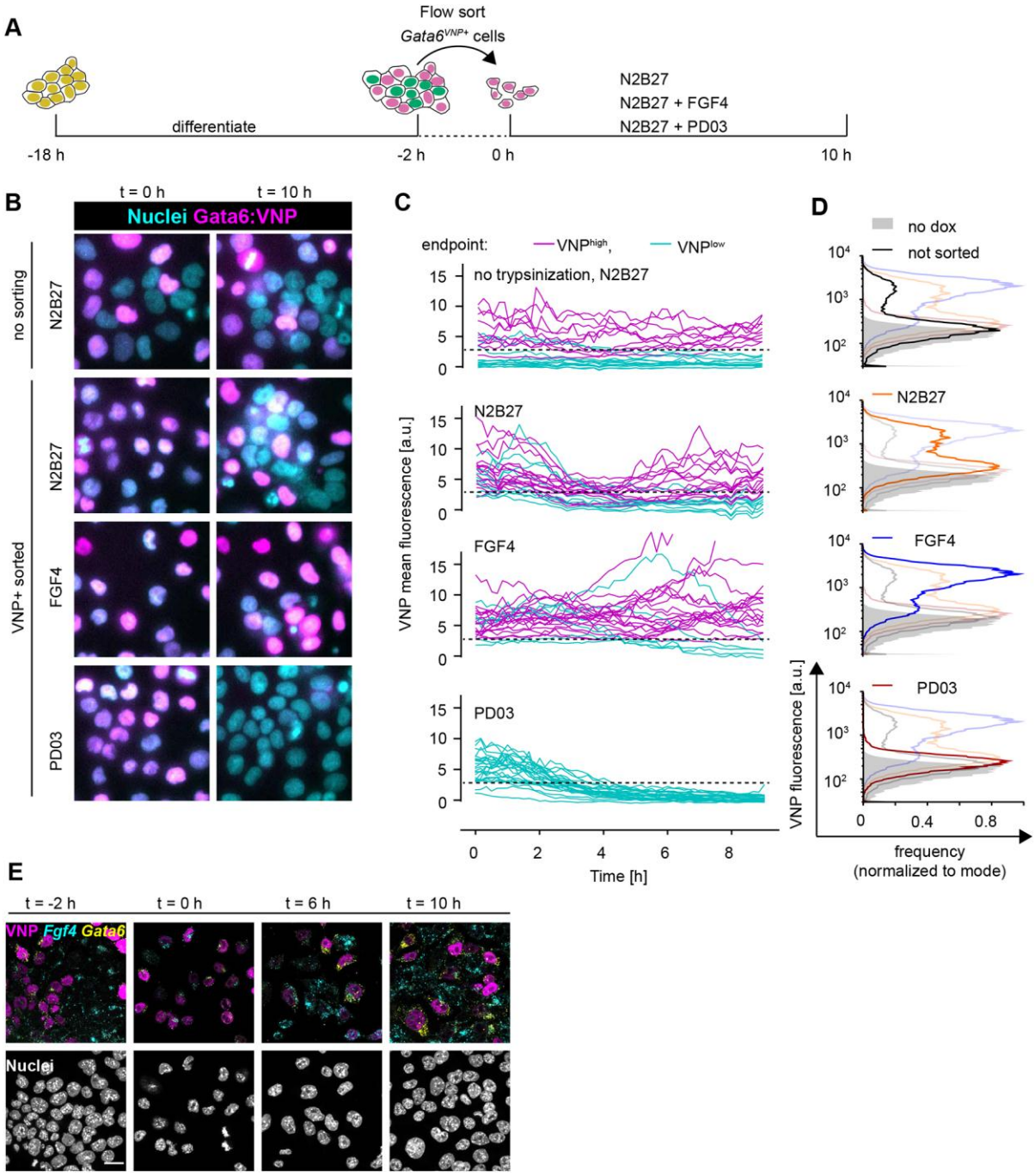


Fig. 6. Heterogeneous cell identities are re-established by cell-cell communication

A Schematic of the cell-sorting experimental protocol. **B** Representative images of Gata6^{VNP} reporter expression in live cells of a non-trypsinised control (upper row) and in cells sorted for VNP expression. Left column is immediately after sorting, right column is after 10 h of culturing in N2B27 medium with the indicated supplements. **C** VNP expression dynamics in individual cells from non-trypsinised colonies (upper panel), or in cells sorted for VNP expression upon culture in the indicated media. Traces are colour coded according to expression levels at the end of the experiment (VNP-high: magenta; VNP-low: cyan). Dashed line indicates the threshold to separate VNP-high from VNP-low cells. **D** Flow cytometry histograms of VNP expression of cells that had not been trypsinised and

Development • Accepted manuscript

sorted (black, top), and of cells that had been sorted for VNP expression followed by 10 h of culture in the indicated media. Each panel shows the histogram of the relevant condition as a dark line, and distributions of all other conditions shaded for comparison. Non-induced control is in grey in all panels. **E** Staining for Fgf4 (cyan) and Gata6 (yellow) mRNA in Gata6^{VNP} reporter cells before sorting (left) and at 2 h, 6 h and 10 h after flow sorting of VNP-positive cells and culture in N2B27. VNP fluorescence in magenta. Scale bar, 20 µm.

Development • Accepted manuscript

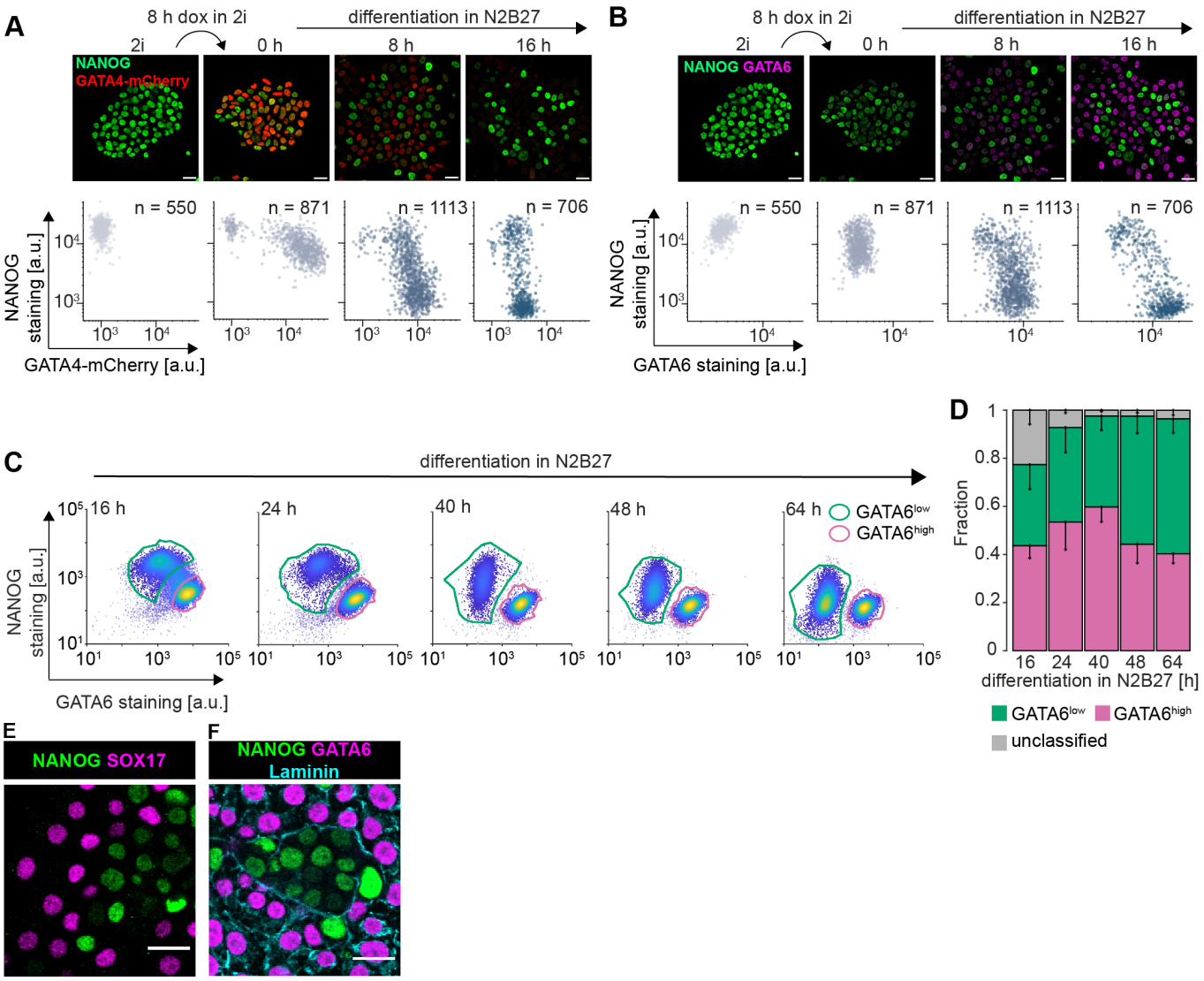


Fig. S1. Differentiation of Epi- and PrE-like cells following pulsed GATA4-mCherry expression

A Immunostaining (top) and single cell quantification (bottom) of GATA4-mCherry (red) and NANOG expression (green) in Gata4-mCherry inducible cells cultured in 2i + LIF medium (left) or at indicated times after the end of an 8 h doxycycline pulse. **B** Same cells as in **A**, but showing immunostaining (top) and single cell quantification (bottom) of GATA6 (magenta) and NANOG (green) expression. **C** Flow cytometry profiles of Gata4-mCherry inducible cells stained for NANOG and GATA6 after an 8 h doxycycline pulse, followed by differentiation in N2B27 medium for the indicated durations. Green and magenta lines surround cells that were assigned to one of the two main clusters by a Gaussian mixture model (methods). **D** Average proportions of cells assigned to one of the two main clusters shown in **C**. Proportion of GATA6^{high} cells in magenta; GATA6^{low} cells in green, and unassigned cells in grey. N = 3, error bars: 95% CI. **E** Immunostaining for SOX17 (magenta) and NANOG (green) in GATA4-mCherry inducible cells cultured for 40 h in N2B27 after an 8 h doxycycline pulse. **F** Immunostaining for GATA6 (magenta), NANOG (green), and Laminin (cyan) in GATA4-mCherry inducible cells cultured for 40 h in N2B27 after an 8 h doxycycline pulse. Scale bars in **A**, **B**, **E**, **F**, 20 µm.

Development • Supplementary information

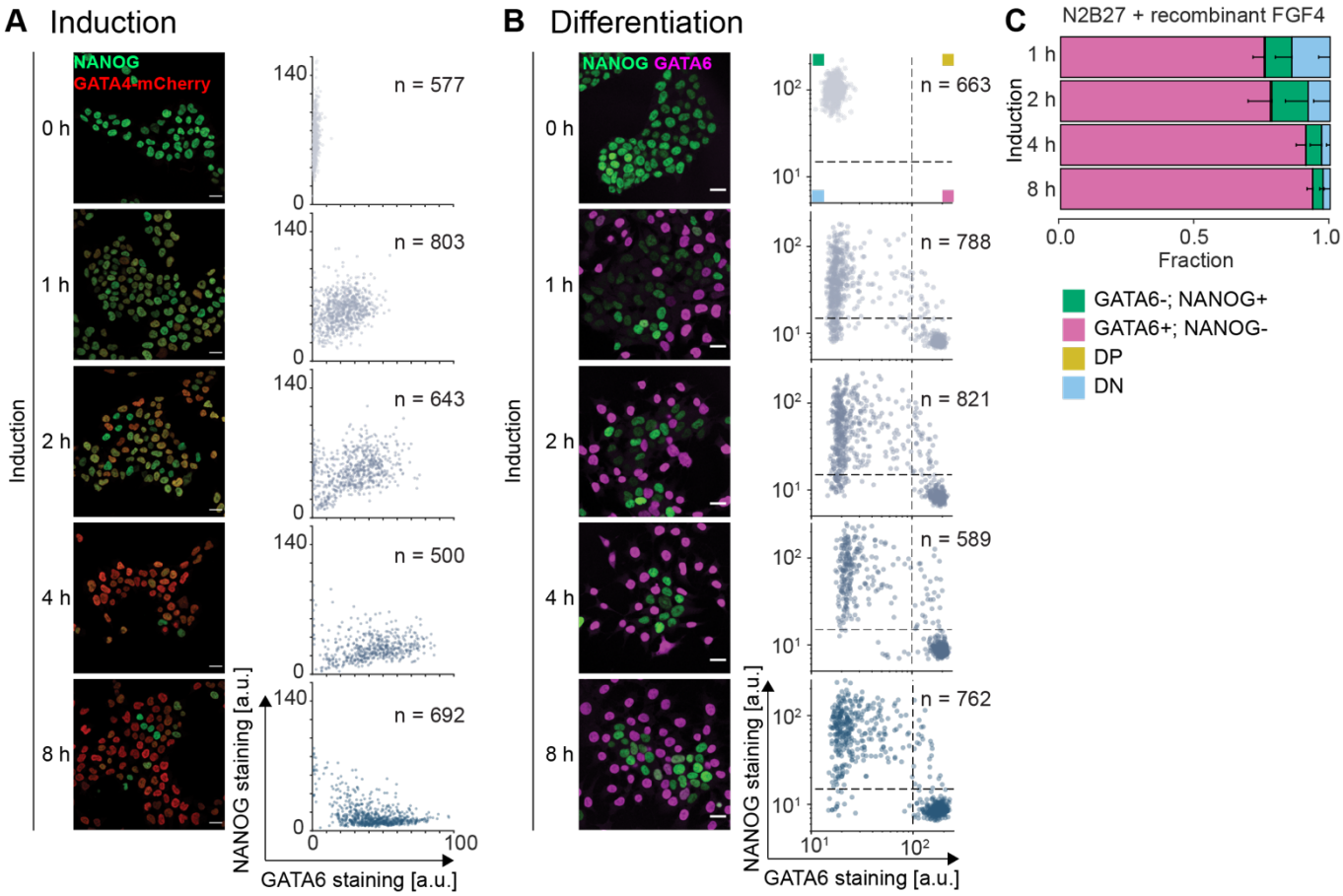


Fig. S2. Quantitative immunofluorescence analysis of NANOG and GATA expression for different doxycycline induction times

A Immunostaining (left) and single-cell quantification (right) of NANOG (green) and GATA4-mCherry (red) expression in inducible cells after indicated durations of doxycycline treatment in 2i + LIF medium. **B** Immunostaining (left) and single-cell quantification (right) of NANOG (green) and GATA6 (magenta) expression in inducible cells after indicated durations of doxycycline treatment, followed by 40 h differentiation in N2B27 medium. Immunofluorescence micrographs in **A** and **B** for 0, 2, and 8 h of doxycycline induction are reproduced from Fig. 1 for comparison. Dashed lines in scatter plots in **B** indicate thresholds to assign cell types; upper left quadrant: GATA6-; NANOG+; lower right quadrant GATA6+; NANOG-; upper right quadrant double positive (DP); lower left quadrant double negative (DN). Scale bars, 20 µm. **C** Proportions of cell types upon indicated durations of doxycycline induction followed by 40 h of differentiation in N2B27 medium supplemented with 10 ng/ml FGF4. Cell identities were determined by immunostaining and quantitative immunofluorescence. Fraction of GATA6+, NANOG- cells in magenta, GATA6-, NANOG+ cells in green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. N = 4, error bars: 95% CI.

Development • Supplementary information

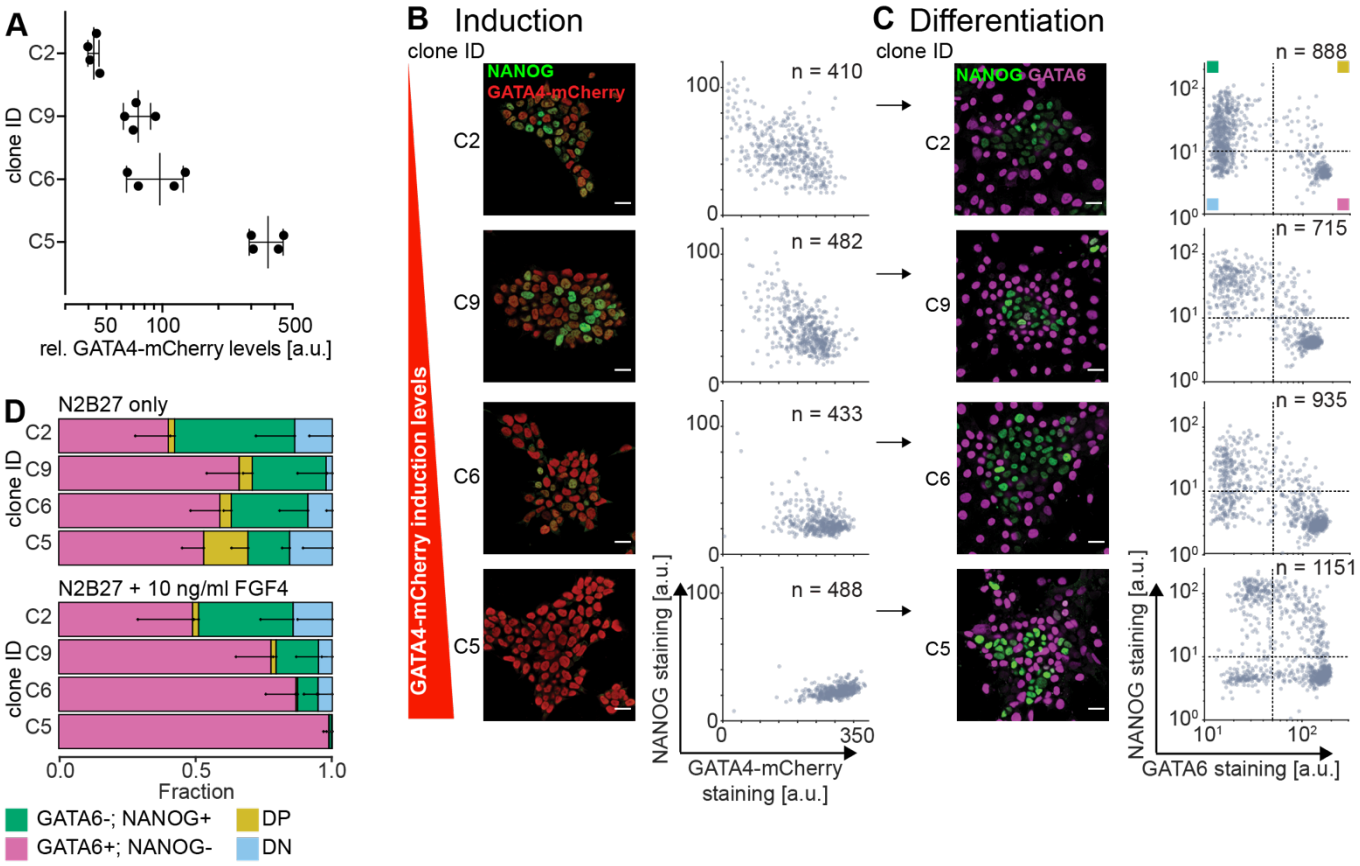


Fig. S3. Robust cell type proportioning in independent clonal cell lines

A GATA4-mCherry induction levels in four independent clonal inducible cell lines after 8 h of doxycycline treatment in 2i + LIF medium, measured by flow cytometry. Clones are ordered by GATA4-mCherry expression strength, fluorescence values were normalized to non-induced control cells. Plot shows individual data points and mean ± SD from N = 4 independent experiments. **B** Immunostaining (left) and single-cell quantification (right) of NANOG (green) and GATA4-mCherry (red) expression in the same clonal lines analysed in **A** after 8 h of doxycycline stimulation in 2i + LIF medium. **C** Immunostaining (left) and single cell quantification (right) of NANOG (green) and GATA6 (magenta) expression in cells from independent clonal lines treated with doxycycline for 8 h and differentiated in N2B27 for 40 h. Dashed lines: Thresholds to determine cell types; upper left quadrant: GATA6-; NANOG+; lower right quadrant GATA6+; NANOG-; upper right quadrant double positive (DP); lower left quadrant double negative (DN). Clones are ordered by GATA4-mCherry induction strength as in **A**. Scale bars in **A**, **C**, 20 µm. **D** Average cell type proportions in clonal lines differentiated as in **C** in N2B27 only (top) or in N2B27 supplemented with 10 ng/ml FGF4 (bottom). Fraction of GATA6+, NANOG- cells in magenta, GATA6-, NANOG+ cells in green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. N = 4, error bars: 95% CI.

Development • Supplementary information

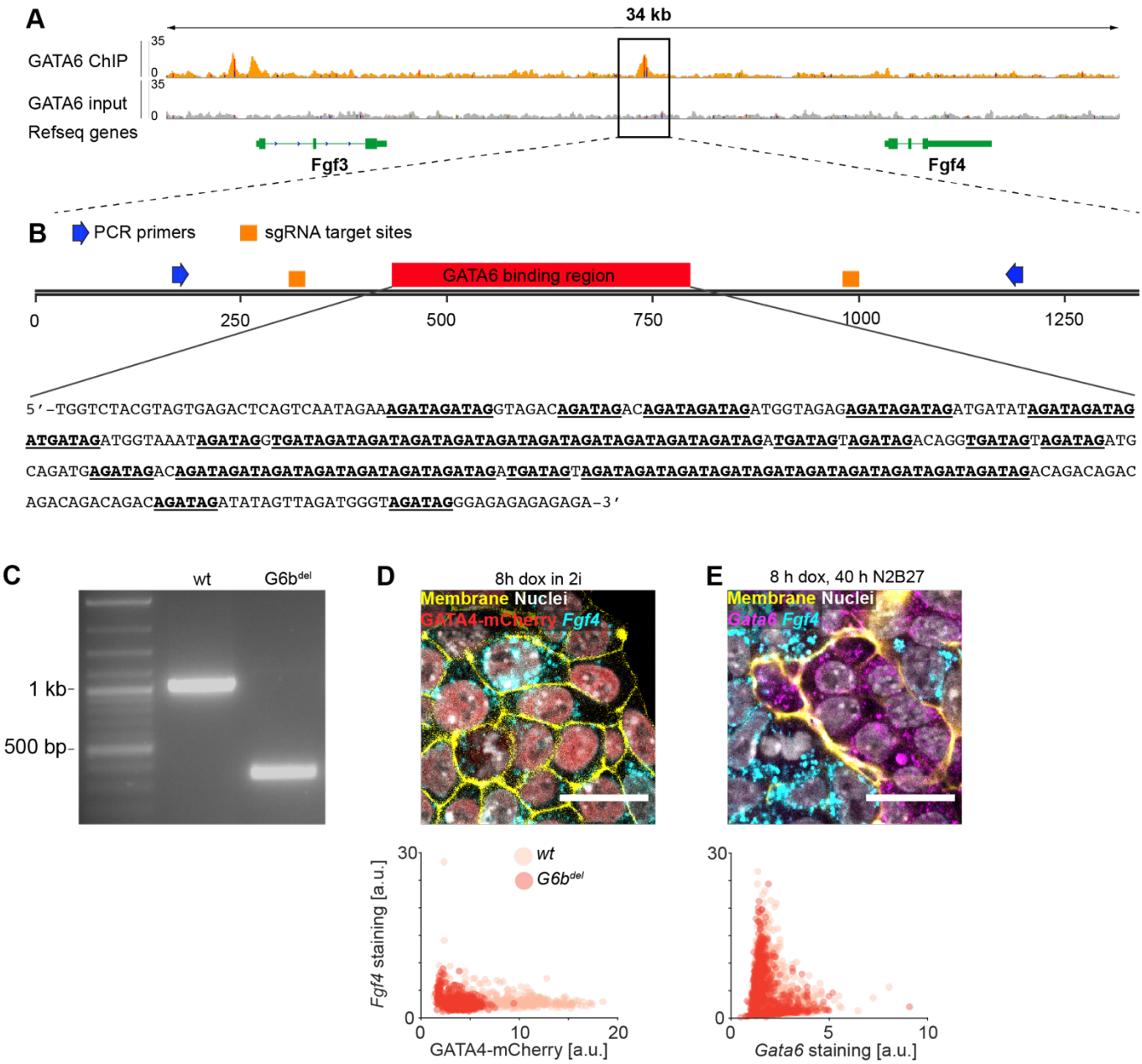


Fig. S4. Characterisation of a GATA6-binding site upstream of Fgf4

A ChIP-seq profile showing GATA6 binding upstream of *Fgf4*. The input control profile is included for comparison (grey). Raw data is from Wamaitha et al., 2015. **B** Schematic overview (top) and nucleotide sequence (bottom) of the GATA6-binding region upstream of Fgf4. Sequences with the GATA6 consensus motif A/T-GATA-A/G identified are underlined and highlighted in bold. **C** PCR validation of successful deletion of the GATA6 binding site in inducible cells. Location sgRNAs used for deletion and PCR primers used for detection are indicated in **B** as orange boxes and blue arrows, respectively. G6b^{del} indicates cell line with a deletion of the GATA6 binding site. **D** Top: GATA4-mCherry expression (red) and Fgf4 mRNA (cyan) in G6b^{del} cells that lack the GATA6 binding site at the end of an 8 h doxycycline pulse. Nuclei stained with Hoechst33342 (white), cell membranes labelled with CellBrite (yellow). Bottom: Corresponding single cell analysis. Data from G6b^{del} cells shown in dark red, same analysis for wild type cells is reproduced from Fig. 2B and shown in light red for comparison. **E** Top: Gata6 (magenta) and Fgf4 mRNA (cyan) expression in G6b^{del} cells after 40 h of culture in N2B27 following an 8 h doxycycline pulse. Nuclei stained with Hoechst33342 (white), cell membranes labelled with CellBrite (yellow). Bottom: Corresponding single cell analysis. Data from G6b^{del} cells in dark red, same analysis for wild type cells reproduced from Fig. 2C shown in light red for comparison. Scale bars in **D**, **E,** 20 µm.

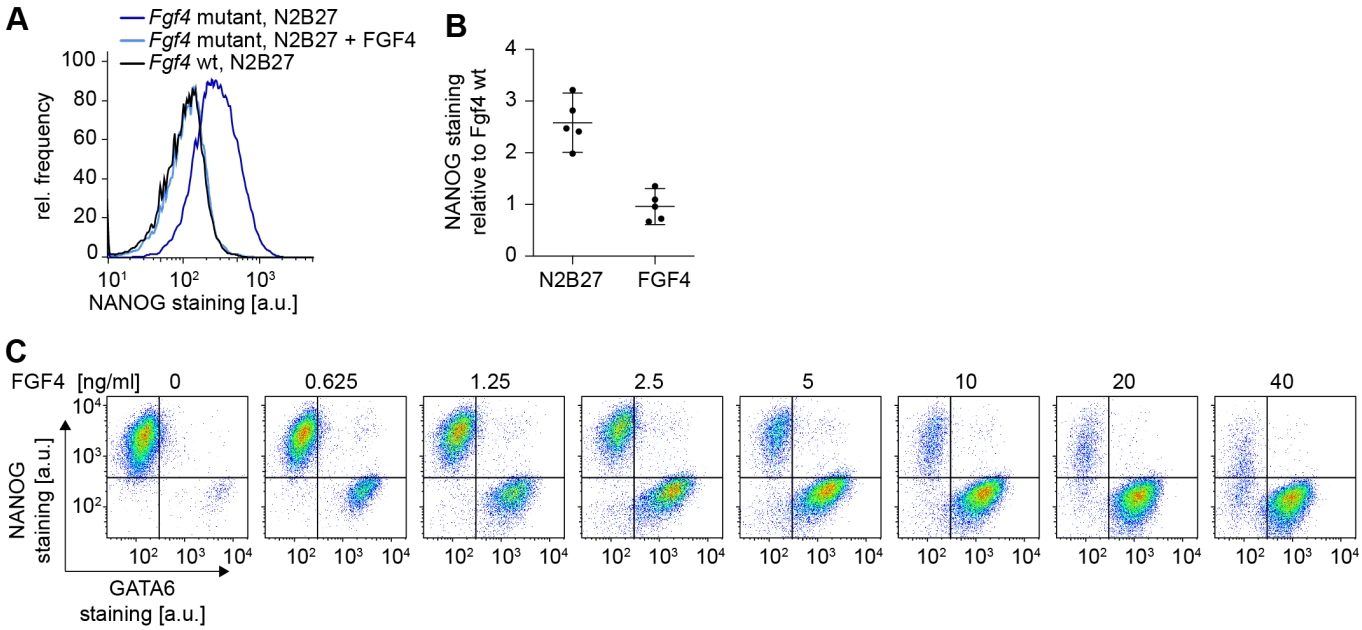


Fig. S5. Recombinant FGF4 rescues differentiation in *Fgf4* mutant cells

A Flow cytometry histograms of NANOG expression in wild type (black) or *Fgf4* mutant (blue) cells after 40 h of culture in N2B27 with (light blue) or without recombinant FGF4 (dark blue, black). **B** Quantitative analysis of NANOG expression levels measured by flow cytometry of *Fgf4* mutant cells cultured for 40 h in N2B27 with or without recombinant FGF4. Dots: Mean fluorescence intensities from ≥ 20.000 cells normalized to NANOG staining levels in the wild type control in individual experiments. N = 4, error bars: 95% CI. **C** Flow cytometry profiles of *Fgf4* mutant cells stained for NANOG and GATA6 after 8 h of doxycycline induction, followed by 40 h of differentiation in N2B27 medium supplemented with 1 µg/ml heparin and the indicated concentrations of recombinant FGF4. Lines indicate gates to assign cell types.

Development • Supplementary information

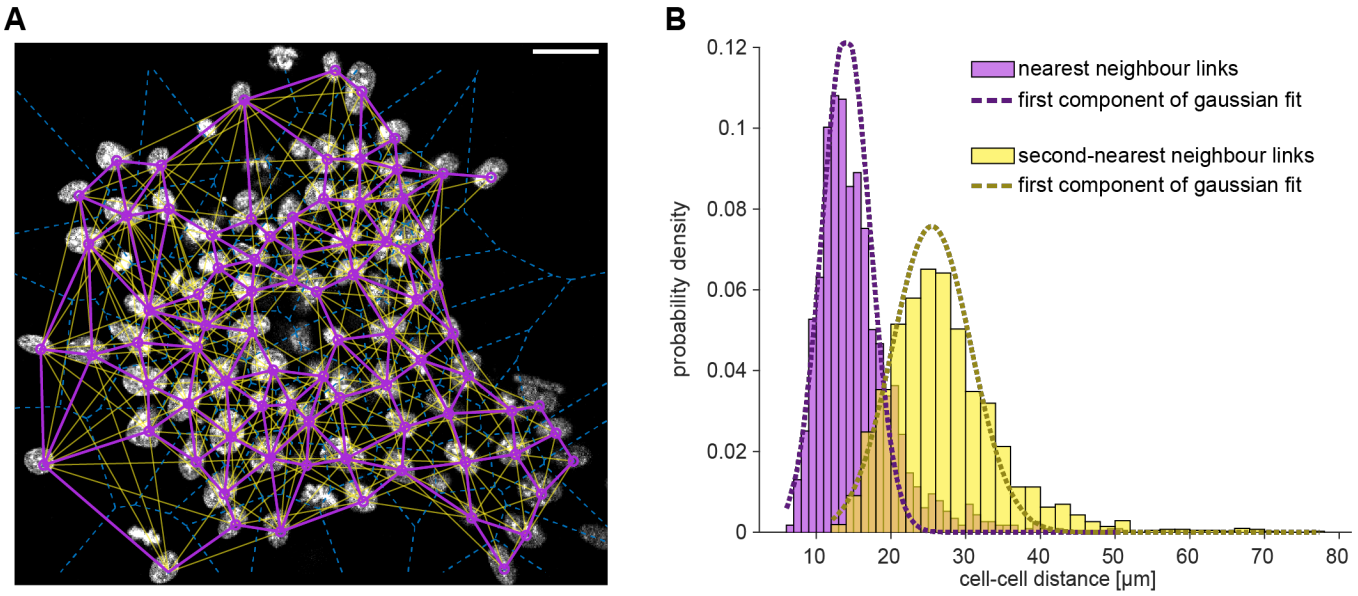


Fig. S6. Estimation of cell-cell distances by Delaunay triangulation

A Nuclear staining (white) of a colony 16 h after initiation of differentiation, overlaid with centres of mass of individual nuclei (purple circles), Voronoi edges (dashed blue lines), and links to nearest (purple) and second nearest neighbours (yellow, see Methods). Scale bar, 20 µm. **B** Histogram of distance distributions between nearest neighbours (purple) and second nearest neighbours (yellow), determined as in **A**. To separate true nearest and second-nearest neighbour distances from spurious longer-distance links, we applied two-component Gaussian mixture model fits to each of the distributions. Dashed lines indicate the first components of each fit. Distances determined by these first components were 14.0 ± 3.2 (mean ± SD) for nearest neighbours, and 25.5 ± 5.3 µm for second nearest neighbours. n = 8 independent fields of view.

Development • Supplementary information

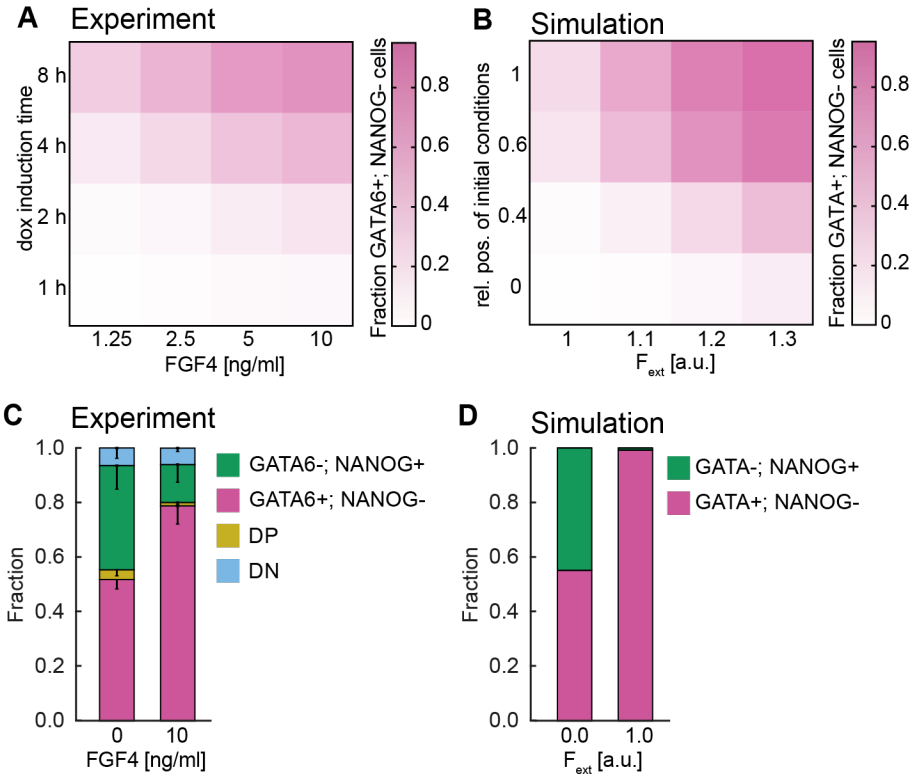


Fig. S7. FGF4-dependence of cell type proportions in *Fgf4* mutant and wild type cells

A Proportion of GATA6+;NANOG- cells in *Fgf4* mutant cultures after 40 h of differentiation in N2B27 as a function of prior doxycycline induction time (vertical axis) and FGF4 concentration during differentiation (horizontal axis). Each square corresponds to one combination of induction time and FGF4 dose, the proportion of GATA6+; NANOG- cells is colour coded in hues of magenta. **B** Proportion of GATA+ cells in simulations of the single cell model for combinations of different initial conditions (vertical axis) and FGF signalling strengths (horizontal axis). Each square corresponds to one combination of initial conditions and FGF signalling strength, the proportion of GATA+ cells is colour coded in hues of magenta. **C** Cell type proportions in wild type GATA4-mCherry inducible cells after an 8 h doxycycline pulse followed by differentiation in N2B27 alone (left) or in N2B27 supplemented with 10 ng/ml FGF4 (right), measured by flow cytometry. Fraction of GATA6+, NANOG- cells in magenta, GATA6-, NANOG+ cells in green, double positive cells (DP) in yellow, and double negative cells (DN) in blue. N = 4, error bars: 95% CI. **D** Cell type proportions in simulations of the coupled model without (left), or with addition of an exogenous FGF signal (right). Initial conditions were chosen such that NANOG and GATA had similar expression levels. Data in **C**, **D** for conditions without exogenous FGF signal is reproduced from Fig. 4C and 4G for comparison.

Development • Supplementary information

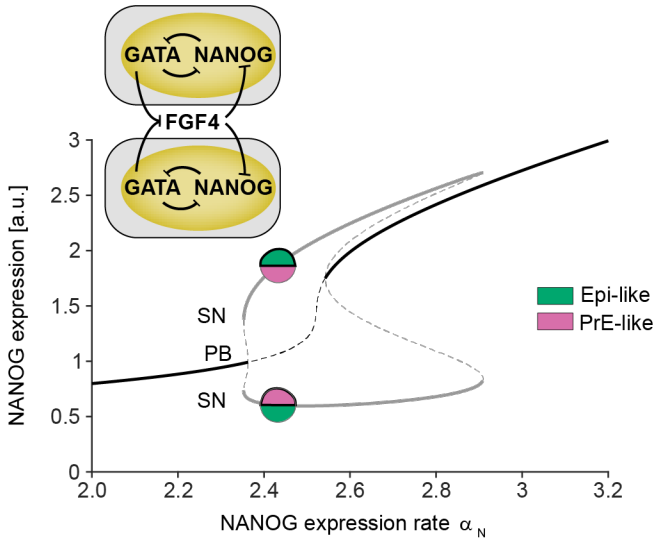


Fig. S8. Bifurcation analysis of a 2-cell coupled model

Bifurcation diagram showing the principle of emergence of a symmetry-broken inhomogeneous steady state (IHSS) for a two-cell coupled system (inset) in dependence of the production rate constant of NANOG (αN), using NANOG expression as a representative variable. The IHSS (grey lines) is generated via a pitchfork (PB) and stabilized via saddle-node (SN) bifurcations. The two branches represent the heterogeneous attractors: (GATA6-/NANOG+) for cell 1 and (GATA6+/NANOG-) for cell 2, or (GATA6+/NANOG-) for cell 1 and (GATA6-/NANOG+) for cell 2 (representative circles). Solid/dashed lines: stable/unstable solutions. Black: homogeneous steady state; grey: IHSS solution.

Development • Supplementary information

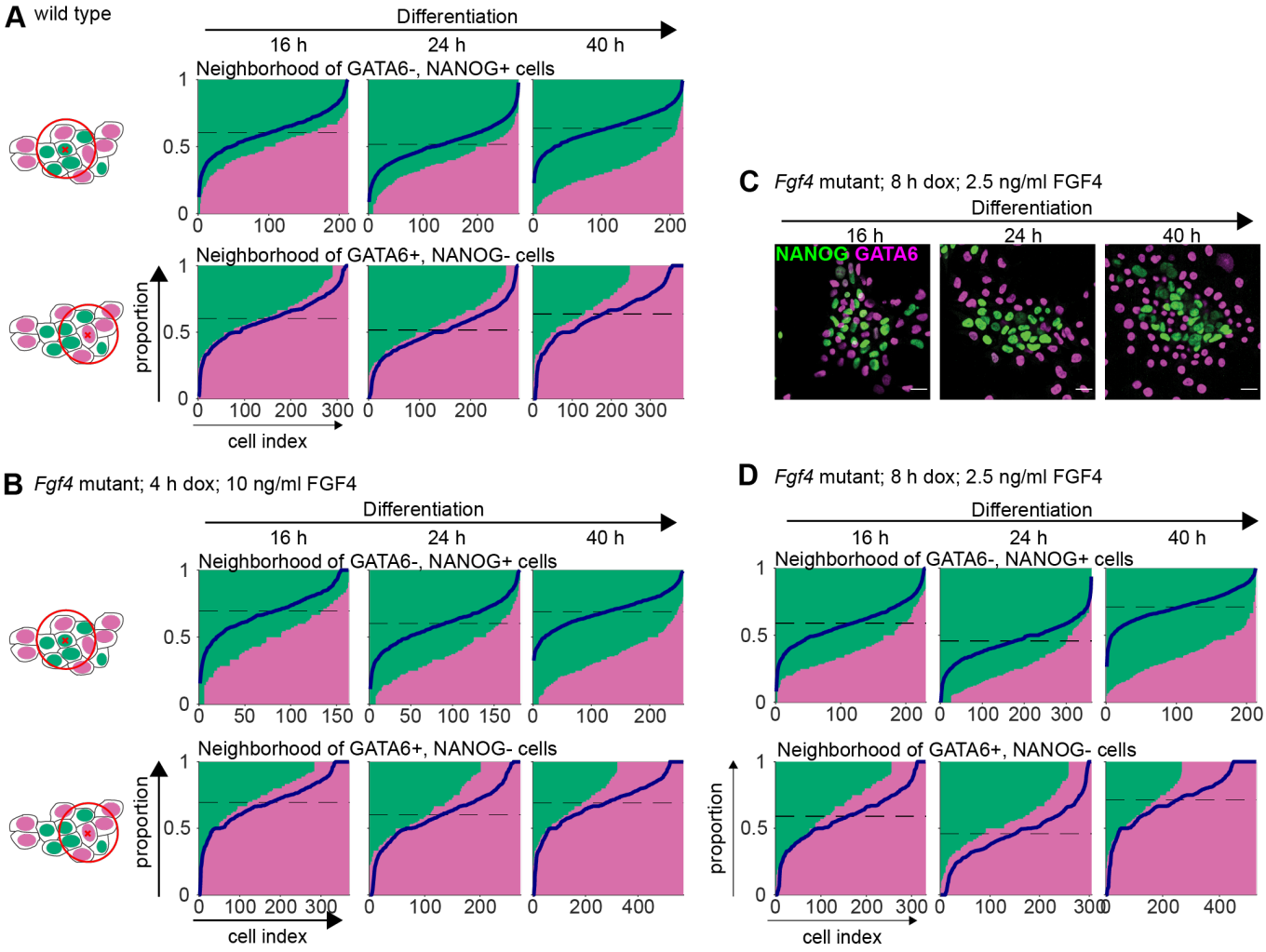


Fig. S9. Cell type composition of local cell neighbourhoods

A Cell type proportions in a local neighbourhood of 31.7 µm diameter in wild type cells stimulated for 8 h with doxycycline and differentiated for the indicated times. Each column in a panel corresponds to one cell, and the sizes of its associated green and magenta bars correspond to the fraction of GATA6-, NANOG+ (green) and GATA6+, NANOG- (magenta) cells in its local neighbourhood. Blue lines indicate the corresponding random distribution, calculated by assigning the global distribution of cell types of the time point randomly to experimentally determined cell positions. Clustering of cell types results in the deviation of the true proportions from this random distribution. The local neighbourhood of GATA6-, NANOG+ cells (top row) indicates clustering of this cell type already at 16 h, which becomes more pronounced over time. Strong clustering of GATA6+, NANOG- cells (lower row) can only be observed at 40 h. **B** Same analysis as in **A**, but for *Fgf4* mutant cells induced for 4 h and differentiated in N2B27 supplemented with 10 ng/ml FGF4 for the indicated periods of time. Deviation from random arrangement (blue lines) indicates clustering of GATA6-, NANOG+ cells at 16 h, which becomes more pronounced over time. Some clustering of GATA6+, NANOG- cells (lower row) can already be detected at 16 h, and becomes more pronounced until 40 h. **C** Immunostaining for NANOG (green) and GATA6 (magenta) in *Fgf4* mutant cells induced with doxycycline for 8 h and differentiated in N2B27 supplemented with 2.5 ng/ml FGF4 for the indicated periods of time. **D** Same analysis as in **A, B**, for *Fgf4* mutant cells differentiated as described in **C**. Deviations from random arrangement (blue lines) are similar to those seen in **B**, indicating similar clustering dynamics.

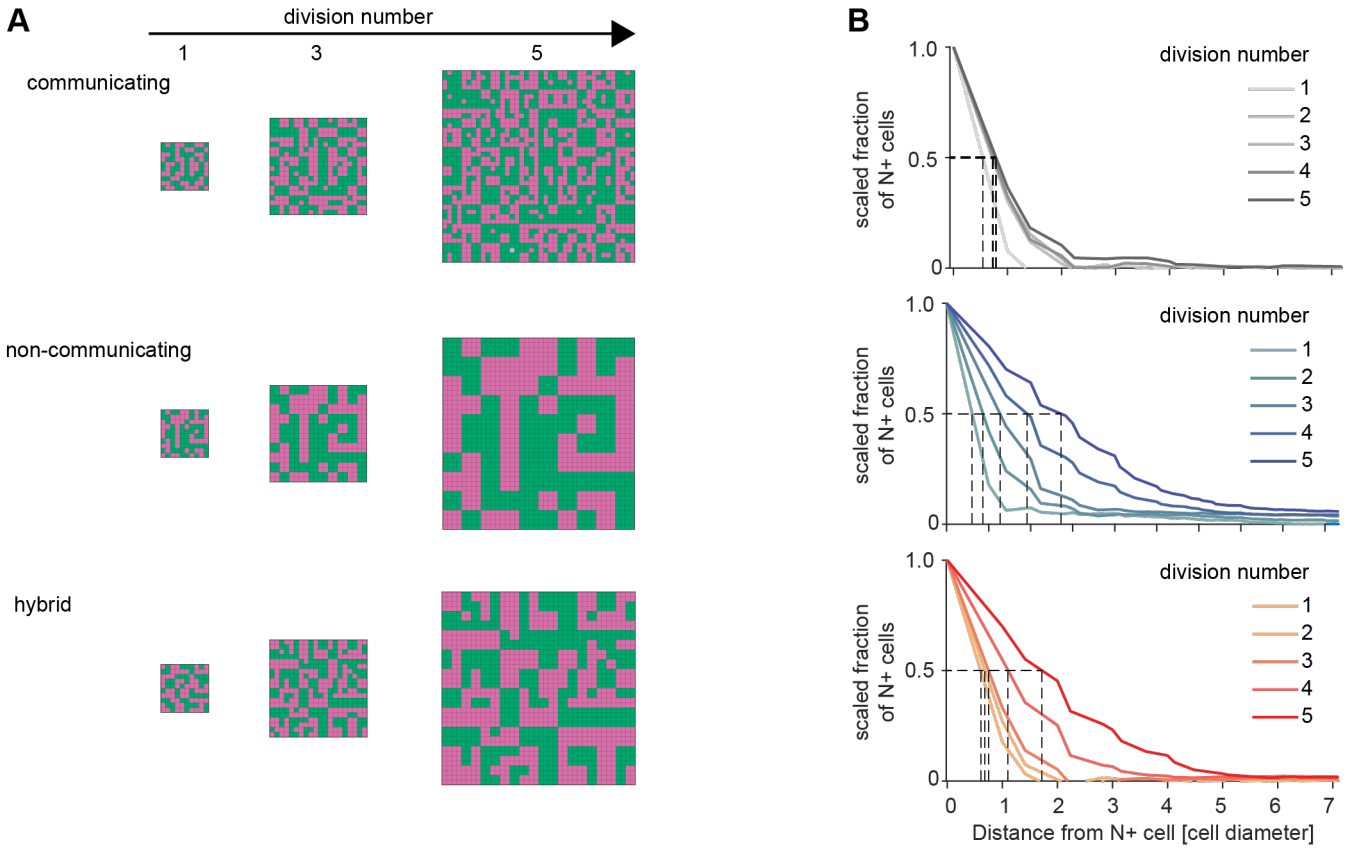


Fig. S10. Spatial patterns of cell types in model simulations

A Spatial configurations of cell types in single numerical realizations of the model with cell-cell communication (top row), without cell-cell communication (middle row) and of a hybrid model (bottom row) where communication is switched off after the third division. First column: Cell type arrangements at the end of the first cell cycle (grid size 10 x 10), second column - end of the third cell cycle (grid size 20 x 20), third column - end of the fifth cell cycle (grid size 40 x 40). **B** Corresponding estimation of NANOG cluster radius for single numerical realizations of the model variants shown in **A**. Graphs depict the scaled fraction of N+ cells within a specific radius around seed cells. Dashed lines indicate determination of cluster radius, the distance around NANOG+ cells at which the scaled fraction is equal to 0.5. Cluster radii were obtained after each cell cycle of the lineage tree simulation.

Development • Supplementary information

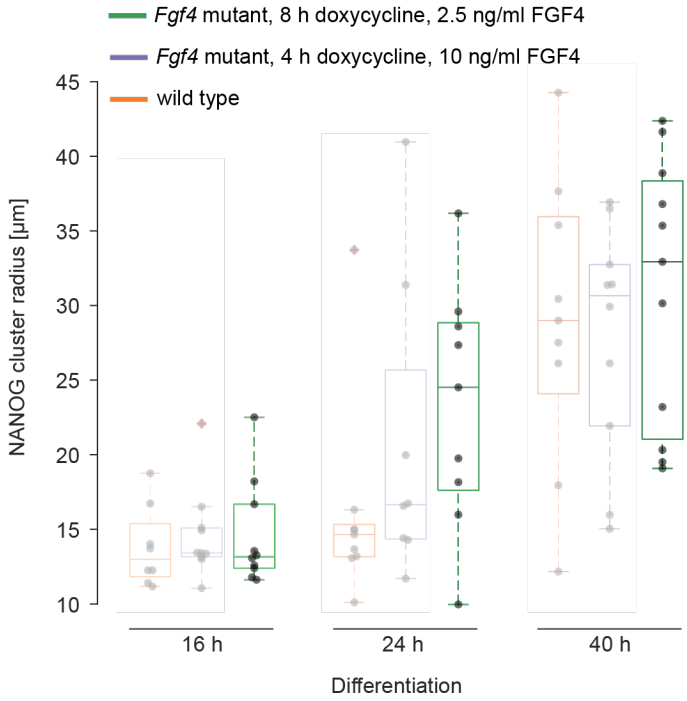


Fig. S11. Cluster radius dynamics in *Fgf4* mutant cells are independent from stimulation and differentiation regime

Dark: Cluster radii of N+ cells in *Fgf4* mutant cultures induced with doxycycline for 8 h, followed by differentiation in N2B27 medium supplemented with 2.5 ng/ml FGF4 for the indicated times. Dots indicate values from individual fields of view (N = 1, n ≥ 9), box plots show median, interquartile ranges and outliers (red cross). Shaded data points are reproduced from Fig. 5C for comparison, and show cluster radii for wild type cells differentiated in N2B27 only (orange), and for *Fgf4* mutant cells differentiated in N2B27 supplemented with 10 ng/ml FGF4 following 4 h of induction with doxycycline (blue).

Development • Supplementary information

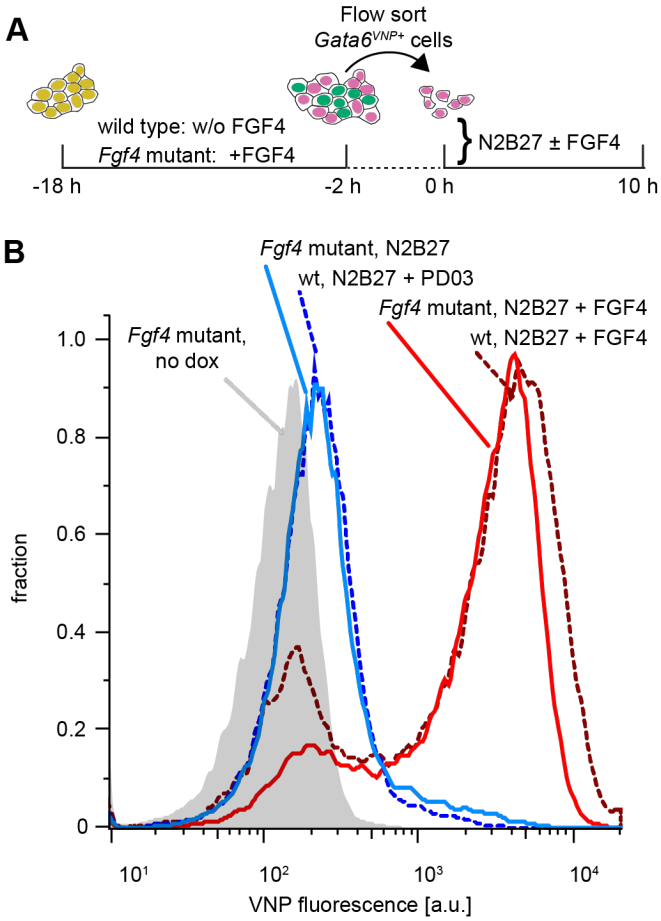
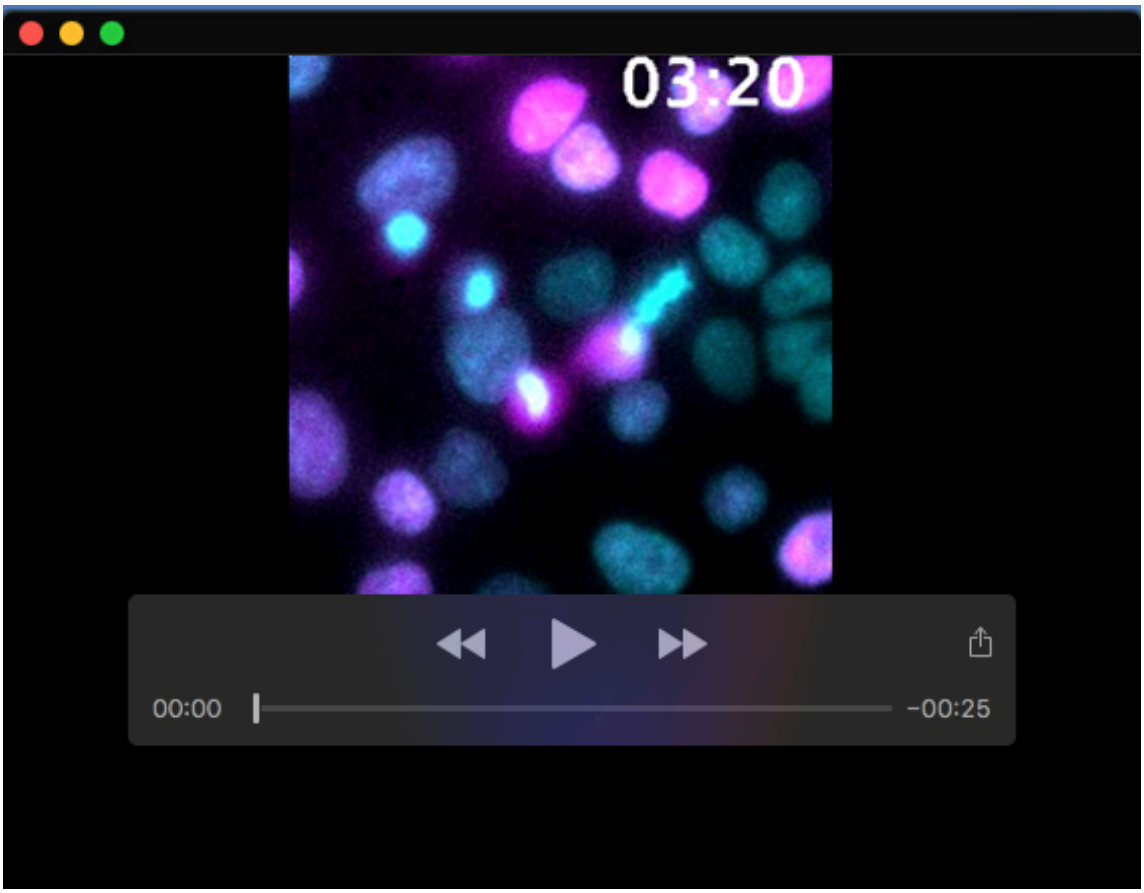


Fig. S12. Gata6^{VNP} reporter expression in sorted *Fgf4* mutant cells

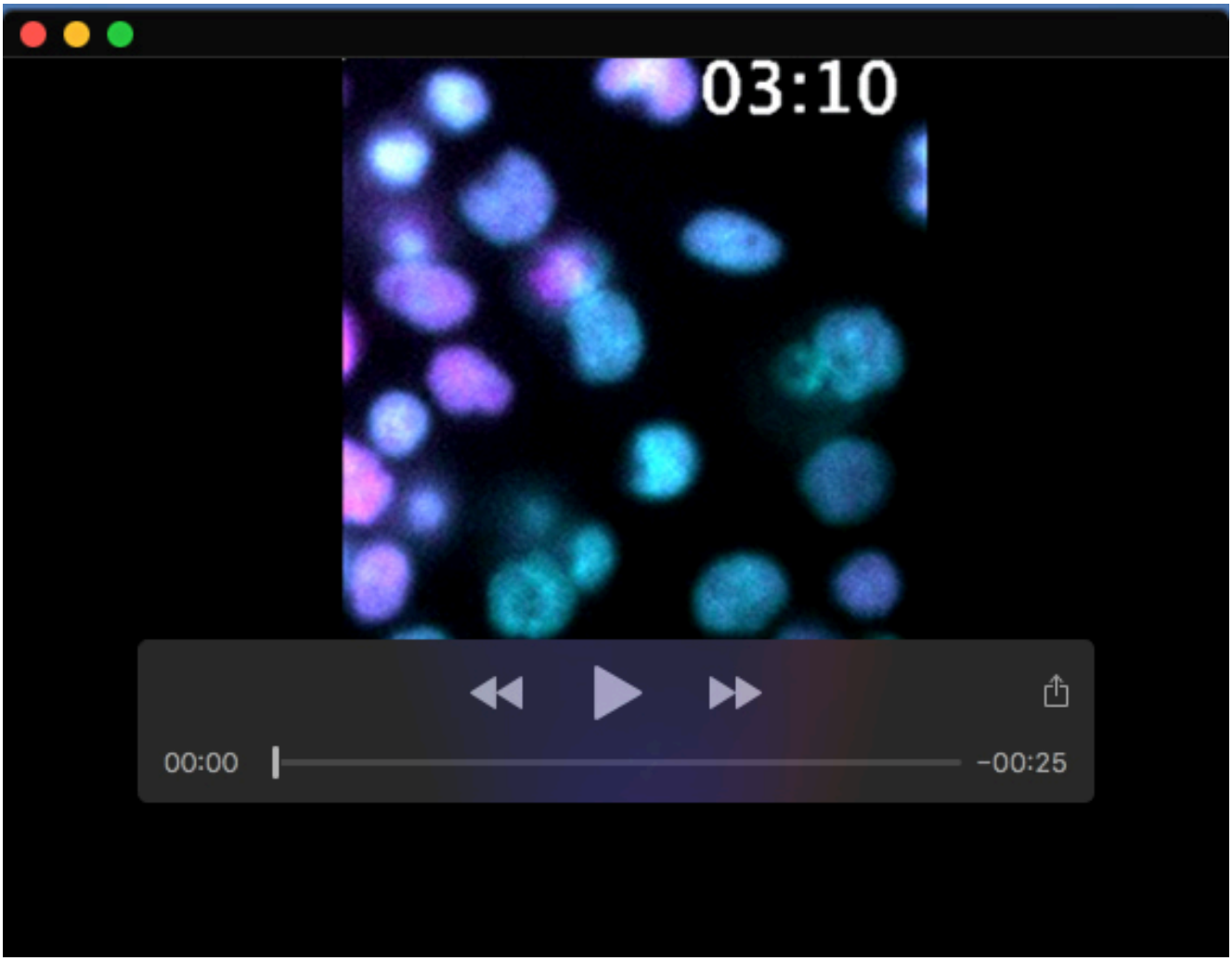
A Schematic representation of the cell-sorting experimental protocol for *Fgf4* mutant cells. Following a doxycycline pulse, *Fgf4* mutant cells were initially differentiated in medium supplemented with FGF4, while wild type cells were differentiated without recombinant FGF4. After flow sorting of VNP positive cells at 16 h, wild type and *Fgf4* mutant cells were cultured in N2B27 medium with or without FGF4, followed by analysis of VNP expression. **B** Flow cytometry histograms of VNP expression in wild type (dashed lines) and *Fgf4* mutant (solid lines) inducible Gata6^{VNP} cells, 10 h after sorting of VNP-positive cells and culture in the indicated media.

Development • Supplementary information



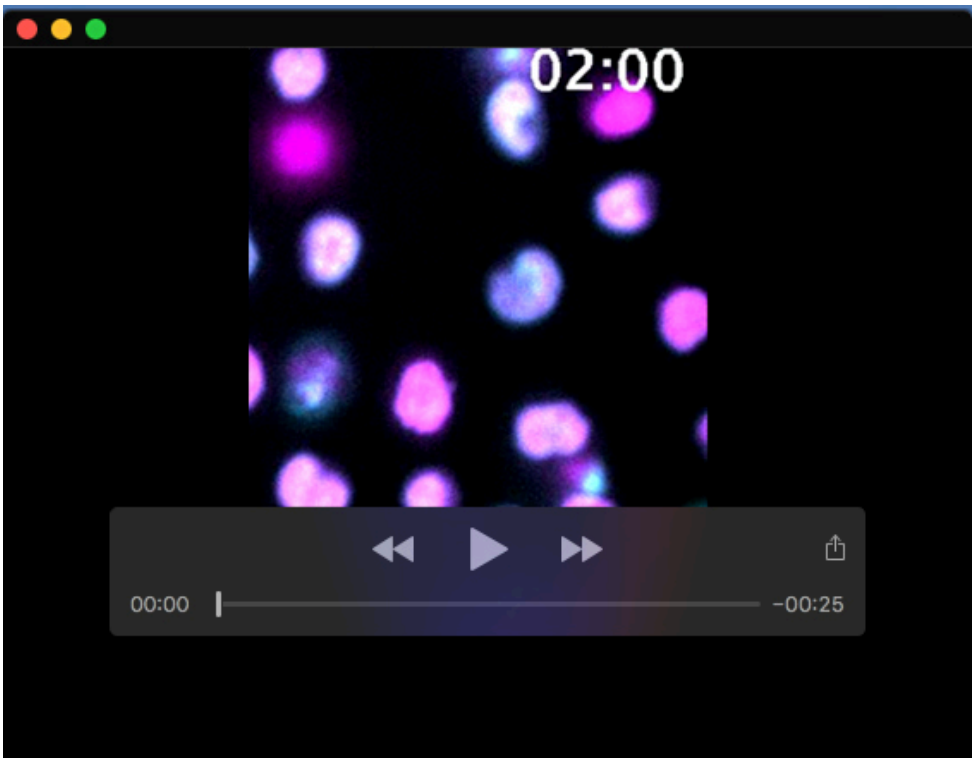
Movie 1. Dynamics of VNP expression in unperturbed colonies
Time-lapse imaging of a colony of Gata6^{VNP} reporter cells starting 16 h after the end of a doxycycline pulse. Medium has been switched to N2B27 at the beginning of the recording.

Development • Supplementary information



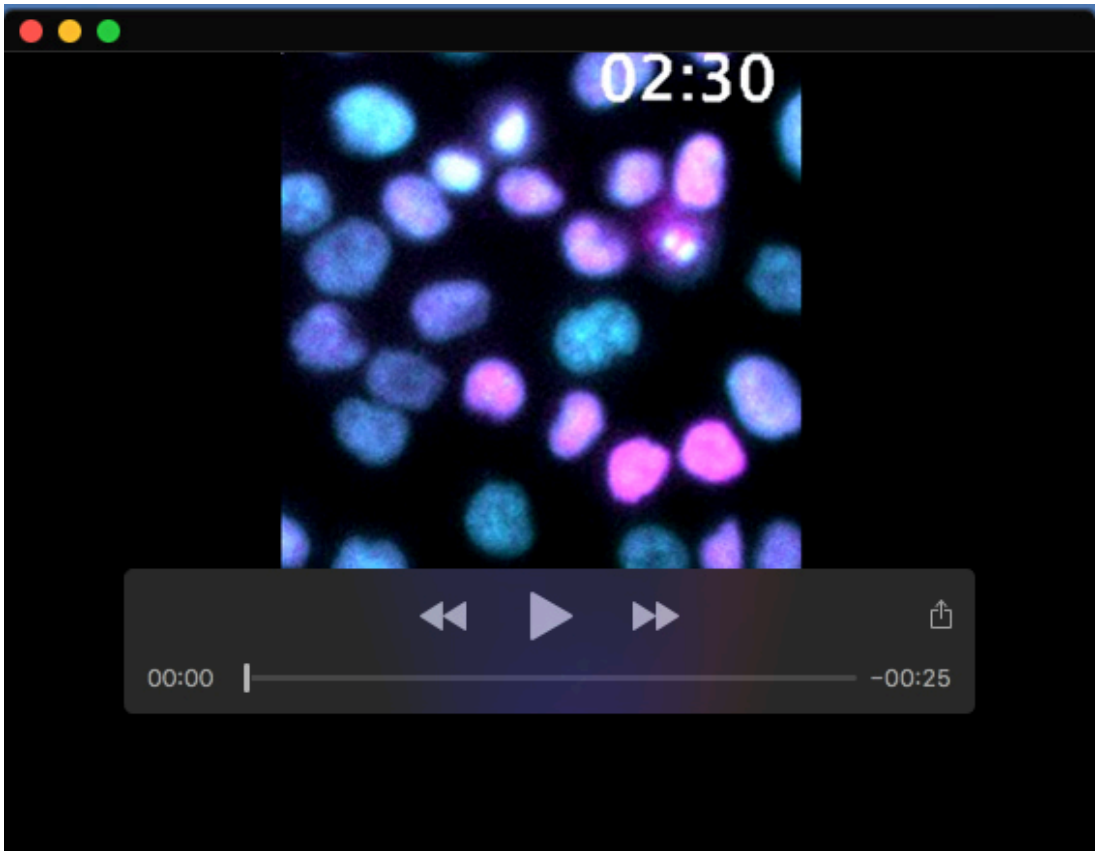
Movie 2. Dynamics of VNP expression in sorted cell populations
Time-lapse imaging of Gata6^{VNP} reporter cells flow sorted for VNP expression 16 h after the end of a doxycycline pulse and cultured in defined N2B27 medium alone (Movie S2), or cultured in N2B27 supplemented with 10 ng/ml FGF4 (Movie S3), or cultured in N2B27 supplemented with 1 μM of the MEK inhibitor PD03 (Movie S4).

Development • Supplementary information



Movie 3. Dynamics of VNP expression in sorted cell populations
Time-lapse imaging of Gata6^{VNP} reporter cells flow sorted for VNP expression 16 h after the end of a doxycycline pulse and cultured in defined N2B27 medium alone (Movie S2), or cultured in N2B27 supplemented with 10 ng/ml FGF4 (Movie S3), or cultured in N2B27 supplemented with 1 μM of the MEK inhibitor PD03 (Movie S4).

Development • Supplementary information



Movie 4. Dynamics of VNP expression in sorted cell populations
Time-lapse imaging of Gata6^{VNP} reporter cells flow sorted for VNP expression 16 h after the end of a doxycycline pulse and cultured in defined N2B27 medium alone (Movie S2), or cultured in N2B27 supplemented with 10 ng/ml FGF4 (Movie S3), or cultured in N2B27 supplemented with 1 μM of the MEK inhibitor PD03 (Movie S4).

Development • Supplementary information