Donor embryonic stem cells displace host cells of 8-cell stage chimeras to the extraembryonic lineages by spatial crowding and FGF4 signalling

Stanley E. Strawbridge[1,2,*,‡], Anna Katharina Schrattel[1,§], Peter Humphreys[1,¶], Kenneth A. Jones[1,**], Jérôme Artus[3,‡‡], Anna-Katerina Hadjantonakis[3], Alexander G. Fletcher[4,5,*], Jennifer Nichols[1,2,*,§§]

[1]Cambridge Stem Cell Institute, University of Cambridge, Jeffrey Cheah Biomedical Centre, Puddicombe Way, Cambridge CB2 0AW, UK
[2]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK
[3]Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[4]School of Mathematical and Physical Sciences, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK
[5]Insigneo Institute, University of Sheffield, Sheffield, S10 2TA, UK

*Authors for correspondence: s.strawbridge@sheffield.ac.uk, a.g.fletcher@sheffield.ac.uk, jenny.nichols@ed.ac.uk
[‡]Present address: Centre for Stem Cell Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK
[§]Present address: Department of Cardiology, Heidelberg University, University Hospital Heidelberg, Im Neuenheimer Feld 410, Germany
[¶]Present address: Cambridge Institute of Science, Altos Labs, Granta Park, Little Abington, Cambridge CB21 6GP, UK
[**]Present address: Department of Biochemistry, University of Cambridge, Sanger Building, Tennis Court Road, Cambridge CB2 1GA, UK
[‡‡]Present address: Université Paris Saclay, Inserm, UMRS1310, 7 rue Guy Moquet, 94800 Villejuif, France
[§§]Present address: MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

Keywords: mouse embryo, epiblast, trophectoderm, primitive endoderm, mathematical modelling, Bayesian inference.

SUMMARY STATEMENT

In mouse blastocyst chimeras, donor cells displace host cells from the EPI by a combination of crowding and FGF4 signalling.

ABSTRACT

Following mouse embryo compaction, outer cells become trophectoderm, while inner cells form the inner cell mass (ICM), later differentiating into primitive endoderm and epiblast during blastocyst formation. Trophectoderm specification is driven by position-governed polarisation, while primitive endoderm specification is positively regulated by FGF4 signalling from the unspecified ICM and epiblast. When injected into an 8-cell stage morula, Embryonic stem cells (ESCs, derived from preimplantation epiblast cells in vitro) can exclude host cells from the epiblast, leading to mice derived entirely from these cells. While evidence suggests roles for ESC-produced FGF4 and physical crowding in host cell displacement from the ICM, the interplay between these possible mechanisms has yet to be dissected, in part due to the lack of studies using *Fgf4*$^{-/-}$ ESCs. Here, we combine chimera titration assays with mathematical modelling to study these mechanisms of host cell displacement. Both *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ ESCs displaced host cells from the epiblast, while only *Fgf4*$^{-/-}$ injected embryos reduced primitive endoderm and increased trophectoderm, indicating sequential exclusion by displacement crowding followed by FGF4 signalling.

INTRODUCTION

The first fate decision in the mouse embryo is driven by positional cues, where outer blastomeres polarise to become the trophectoderm (TE), founding tissue of the placenta (**Fig. 1A, B**). The TE forms an epithelium surrounding the bipotent inner cell mass (ICM) (Hillman, Sherman, and Graham 1972; Tarkowski and Wróblewska 1967), which then segregates into the primitive endoderm (PrE), source of the yolk sac, and epiblast (EPI), future fetus and source of embryonic stem cells (ESCs) (Gardner and Rossant 1979; Martin 1981; Evans and Kaufman 1981; Boroviak and Nichols 2014). Specification of the ICM occurs asynchronously (Plusa et al. 2008; Saiz et al. 2016), largely due to fibroblast growth factor 4 (FGF4) secretion from unspecified ICM cells and EPI (Feldman et al. 1995; Frankenberg et al. 2011; Kang et al. 2013; Nichols et al. 2009; Yamanaka, Lanner, and Rossant 2010). Basement membrane components including Laminin 511 increase the efficiency of ESC capture (Boroviak et al. 2014), and PrE is known to produce basement membrane components like Collagen 4a1 and Laminin 1 (Kang,

Garg, and Hadjantonakis 2017), thus implicating PrE-produced basement membrane in EPI establishment and/or maintenance.

The dynamic allocation of the three lineages can be perturbed in myriad ways, such as re-incorporating ESCs into preimplantation embryos by injection or aggregation to form chimeras (Poueymirou et al. 2007; Grabarek et al. 2012; Humięcka et al. 2016). These donor cells can re-enter normal development from the preimplantation EPI stage and can contribute to all adult germ layers and the germ line (Bradley et al. 1984). Injecting donor ESCs into 8-cell embryos at embryonic day (E) 2.5 can increase TE cell numbers by displacing host blastomeres outward (**Fig. 1B**) (Humięcka et al. 2016). Donor cells can also increase host-derived PrE cell numbers. With sufficient donor cells, the resulting mouse may be entirely donor-derived; in contrast, blastocyst-stage injection yields chimeras that are only partially derived from ESCs (Poueymirou et al. 2007). This modulation in the second fate decision likely stems from FGF4 production by donor ESCs. Indeed, high exogenous FGF4 levels can drive the entire specifying ICM to the PrE fate (Yamanaka, Lanner, and Rossant 2010). While FGF4 is a key driver of ICM specification, no feedback role has been proposed for PrE. However, laser ablation studies show that the specifying ICM will compensate for EPI or PrE loss (Saiz et al. 2020), suggesting population-level feedback between these three cell types.

Here we unite these observations into a theoretical framework. We first generated a compartment model of cell population dynamics in the E2.5-E4.5 mouse embryo, calibrated against previous observations (Saiz et al. 2016), indicating a role for PrE feedback on ICM specification. Using this model, we conducted donor cell injections into host embryos using wild type (WT) *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ ESCs. Both donor types impeded host EPI contribution, with a smaller PrE and larger TE observed in the *Fgf4*$^{-/-}$ case. Finally, we combined our base model with chimera assays to model chimera formation, indicating that donor cells perturb host cell allocation via spatial crowding and subsequent FGF4 induction.

RESULTS AND DISCUSSION

A compartment model of blastocyst generation suggests a role for feedback from the PrE on ICM specification

Previous models of blastocyst generation have focused on position for the first fate-decision and a bistable gene regulatory network for ICM specification (Bessonnard et al. 2014; Nissen et al. 2017; Saiz et al. 2020). Here we aimed to create a minimal model of blastocyst formation,

extendable to generate *in silico* chimeras by adding donor ESCs. This model provides outputs for transition rates and numbers and proportions of cells in each lineage. We designed a compartment model of mouse embryogenesis spanning the E2.5 8-cell stage to the E4.5 late blastocyst stage (**Fig. 1B**). During this period, two binary cell-fate decisions occur. First, blastomeres, B, specify, at a rate □, into unspecified ICM cells, C, with a bias of □, or TE, T, with a bias of $1-\boldsymbol{\rho}$. Second, unspecified ICM cells become either PrE, P, or EPI, E. PrE specification is driven by FGF4 secreted from the unspecified ICM cells and EPI. This is reflected in the PrE specification rate as $\eta\ (C+E)^m$, where η is a constant and m is a feedback parameter to allow for potential non-linearity. For EPI specification, we compare two models. The first is a constant rate of specification, ζ. This emulates one school of thought that the unspecified ICM will take on the EPI identity by default in the absence of FGF4. Our alternative model is an undetermined feedback from the PrE, possibly ECM components such as laminin. In this model the specification rate is $\zeta\ P^l$, where l is a feedback parameter to allow for potential non-linearity. Finally, for simplicity we assumed that all cells proliferate with the same net growth rate □, which captures both cell division and death (**Fig. 1C**). Partial evidence in support of this simplifying assumption is provided by a good linear fit between TE and non-TE cell numbers in our data (**Fig. S1A**).

We inferred parameters for both models using an Approximate Bayesian Computation with Markov Chain Monte Carlo (ABC-MCMC), with data from (Saiz et al. 2016) (**Fig. 1D, Fig. S1B,C**). Both models yield net growth rates around 0.06 h^{-1}, corresponding to a doubling time of ~11 h. This rate allows an 8-cell morula to undergo four doublings, producing ~128 cells. Both models also capture the TE (65%) to unspecified ICM (35%) ratio, with an unspecified ICM bias $\boldsymbol{\rho}=0.35$. To assess the sufficiency of no, linear, or nonlinear feedback, parameters l and m were varied among 0, 1, and 2. In all cases, the inference algorithm selected a nonlinear feedback (value of 2). Both models reproduce TE and PrE proportions well (**Fig. 1E-H**), but the model without PrE feedback fails to capture the dynamics of unspecified ICM and EPI. Since the ICM both generates and responds to FGF4, we hypothesise that PrE specifies earlier than EPI. Supporting this, (Saiz et al. 2016) show more early PrE than EPI cells. However, (Plusa et al. 2008) and (Grabarek et al. 2012) report that EPI plasticity is lost before PrE. A Bayes factor of ~17 (based on 500 acceptances from 6,289 and 105,027 parameter sets respectively, with ε = 17,300) provides positive evidence in favour of the PrE feedback model over the non-feedback model (Toni et al. 2009). This is especially pronounced when considering unspecified ICM, which should be fully resolved into EPI or PrE by E4.5 (**Fig. 1D,G,H**). To capture intrinsic

stochasticity in small cell populations, we ran Gillespie simulations using the median posterior parameter values from the PrE feedback model (**Fig. S1D**). These reproduced observed cell type distributions, validating parameter estimates and confirming that the deterministic model approximates cell population dynamics well.

This suggests that PrE feedback is sufficient to drive timely ICM specification. While we favour the hypothesis that induction of ICM to EPI is mediated by basement membrane components from the PrE, this remains unproven. Further experimental and theoretical work is needed to clarify the mechanism. Next, we expanded on our base model to examine how donor cells perturb host cell allocation, generating two datasets to explore host-donor interactions and tissue modulation in response to donor cells.

Both *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ donor ESCs can impede host cells from contributing to the EPI

Previous work has shown that donor ESCs have the ability to displace host cells from EPI through FGF4 signalling (Poueymirou et al. 2007; Humięcka et al. 2016). To disentangle FGF4 signalling from any other mechanism, we investigated host EPI displacement by injecting either 10 *Fgf4*$^{+/+}$ or 10 *Fgf4*$^{-/-}$ ESCs into wild-type (WT) 8-cell stage morulae, via perforation of intact zona pellucida (**Fig. 2A**). Chimeric embryos were cultured *ex utero,* alongside WT non-injected embryos, for 48 hours to the late blastocyst stage. Embryos were then fixed, immunostained, and imaged for EPI marker SOX2, PrE marker GATA4, and donor cell marker DsRed (**Fig. 2B**). ICM cell numbers and types were determined manually with the ImageJ plugin “Cell Counter” (**Fig. 2C, D**). Two groups with varying levels of chimerism were observed in the 10 *Fgf4*$^{-/-}$ ESC injected (10-/-) embryos. This classification was supported by k-means clustering, which yielded an optimal silhouette score of 0.6252, indicating moderate structure consistent with two biologically distinct subpopulations (**Fig. S2A-C**) (Kaufman and Rousseeuw 1990). The two resulting groups were chimeras with high (Hi) and low (Lo) levels of donor cell contribution to the total EPI.

All three groups of chimeras had EPIs with greater cell numbers than non-injected embryos (**Fig. 2E**). However, when examining the EPI fraction of the ICM, only 10-/- Hi and Lo chimeras had larger EPI compartments. The reason for this became apparent upon inspecting the number of PrE cells within the different chimera conditions (**Fig. 2F**). Both the 10-/- Hi and 10-/- Lo chimeras showed a reduced PrE cell number and fraction of the ICM, while *Fgf4*$^{+/+}$ ESC injected (10+/+) chimeras showed a reciprocal increase in both PrE cell number and fraction.

Finally, 10+/+ and 10-/- Hi chimeras had fewer host cells in their EPI as compared to the 10-/- Lo chimeras and non-injected embryos (**Fig. 2G**).

This suggests alternative mechanisms of host EPI exclusion by *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ donor cells. In 10+/+ chimeras, we propose that FGF4+/+ donor ESCs drive host unspecified ICM cells into the PrE via sustained FGF4 signalling, effectively excluding host cells from the EPI. In contrast, *Fgf4*$^{-/-}$ donor cells lack FGF4 production, implying a different exclusion mechanism. Are potential host EPI cells lost through reduced proliferation or apoptosis, or are they, as in 10+/+ chimeras, redirected to another lineage such as TE? To investigate, we next quantified cell number and lineage identity in late blastocysts.

Embryos injected with *Fgf4*$^{-/-}$ donor ESCs have fewer PrE and more TE cells

We investigated how *Fgf4*$^{-/-}$ donor ESCs perturb host cell allocation by quantifying cell numbers of all three lineages at the late blastocyst stage (**Fig. 3A**). We injected 10 or 15 *Fgf4*$^{+/+}$ (15+/+) or *Fgf4*$^{-/-}$ (15-/-) ESCs into WT 8-cell stage morulae. Chimeric and non-injected WT embryos were cultured for 48 hours to the late blastocyst stage, then fixed, stained for DAPI (nuclei), SOX2 (EPI), and GATA4 (PrE), and imaged (**Fig. 3B**). Cells were segmented and cell types determined as described in *Materials and Methods* (**Fig. 3C-F**).

Consistent with the previous experiment, all chimera conditions showed increased EPI cell number and ICM fraction compared to non-injected embryos (**Fig. 3G**), with EPI scaling proportionally to the number of donor cells, regardless of type. Conversely, PrE cell number and ICM fraction were reduced in 10-/- and 15-/- chimeras (**Fig. 3H**), consistent with prior findings that *Fgf4*$^{-/-}$ donor ESCs impede PrE specification. Compared to earlier 10+/+ chimeras, both 10+/+ and 15+/+ chimeras showed reduced PrE fractions, with 15+/+ exhibiting a greater decrease. Overall, higher numbers of donor cells, regardless of genotype, consistently lowered the PrE fraction within the ICM.

In *Fgf4*$^{-/-}$ ESC-injected embryos, TE cell numbers increased reciprocally with a decrease in PrE cells (**Fig. 3I**). The proportions of 15+/+ and 15-/- chimeras composed of TE cells were significantly smaller than in non-injected embryos, likely due to increased EPI cell numbers. These findings suggest that donor cells, regardless of genotype, can drive host cells into the TE compartment during the first cell-fate decision. In chimeras with *Fgf4*$^{-/-}$ donor cells, remaining host ICM cells can specify into either EPI or PrE. In contrast, in chimeras with *Fgf4*$^{+/+}$ donor

cells, these unspecified host ICM cells are biased toward PrE, likely via FGF4 signalling. We next formalize these concepts in a mathematical model of chimera formation.

Quantitative modelling suggests a role for displacement crowding and FGF4 signalling in host EPI exclusion

We integrated donor cells into our base model of embryogenesis to generate *in silico* chimeras and compared three models of chimera formation. The first model includes FGF4 induction from *Fgf4*$^{+/+}$ donor cells, D^{+}, (**Fig. 4A**) and both *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ donor cells, D^{-}, have the same net growth rate as host cells (F) (**Fig. 4Bi**). The second model includes FGF4 induction and a different net growth rate, differential growth (GF) (**Fig. 4Bii**), and the third model includes FGF4 induction, differential growth, and spatial crowding,which drives host towards the TE, from both types of donor cells (GFC) (**Fig. 4C**). These models were simulated using the posterior median values for parameters from the base model of embryogenesis, including the host cell net growth rate (**Fig. 1B, C**). Additional parameters for the GF and GFC models were estimated by ABC-MCMC using data from **Fig. 3**. Donor cell net growth rate, α_D, was determined for the GF model and donor cell net growth rate, crowding factor, *a*, and feedback parameter, n, were estimated for the GFC model. Models were simulated using the median of the posterior parameter distribution (**Fig. 4D-F, Fig. S3A, B**). The inferred donor growth rate (α_D) was approximately 0.02 h^{-1}, consistent with empirical observations of ~0.018 h^{-1} based on live-cell tracking of cells derived from donor ESCs in chimeric embryos (Alexandrova et al. 2016).

Upon examining model performance, we saw the F model had EPI cell numbers more than five times larger than observed (**Fig. 4D**), while both GF and GFC models better matched EPI cell numbers. However, the GF model performed poorly against 10+/+ and 15+/+ chimeras with low EPI fractions of the ICM. Neither F nor GF models performed well against the PrE cell numbers, overestimating the values in all observed chimeras. The GFC model, on the other hand, follows the PrE trends in both cell number and fraction of the ICM. When considering the TE, the F and GF models show no modulation in TE cell number while the GFC model does (**Fig. 4E**). Finally, we test the predictive power for these models against data they have not been exposed to (**Fig. 4F**). We see that only the GFC model predicts both host EPI cell numbers and fraction of the ICM. Neither the F nor GF models are able to predict host EPI exclusion in terms of cell number for the 10-/- chimeras. The Bayes factor between the GFC model and FC model is around 45 (based on 500 acceptances from 585 and 26,575 parameter sets respectively, using a rejection ABC threshold of ε = 117,000), providing strong evidence in favour of the GFC model over the

FC model (Toni et al. 2009). To further assess how well the GFC model reflects biological variability in small cell populations, we performed stochastic simulations using the median posterior parameter values (**Fig. S3C**). These simulations reproduced both the range and variability of lineage contributions seen in the experimental data, suggesting that the model captures key sources of noise and heterogeneity in chimera development. This suggests that the donor cells perturb host tissues through spatial crowding in the first cell-fate decision and FGF4 induction in the second cell-fate decision (**Fig. 4G**).

It has been shown that the ESCs sort to the interior of the E2.5 8-cell stage embryo, while blastomeres remain on the exterior (Humięcka et al. 2016). The mechanisms underlying this process are not yet established, but could include differential cell contractility (Maître et al. 2016) or membrane fluctuations (Yanagida et al. 2022). As a result, some host cells that would otherwise be located to the interior may be forced to the exterior and specified as TE. For *Fgf4*$^{-/-}$ donor ESCs, the remaining host ICM cells specify normally as either EPI or PrE, whereas for *Fgf4*$^{+/+}$ donor ESCs, host cells show a bias toward PrE. Taken together, these findings advance our understanding of how donor ESCs influence lineage allocation of host cells as ESCs reincorporate into normal development, shedding light on mechanisms of cell-fate plasticity in the early mouse embryo.

MATERIALS AND METHODS

ESC culture

Fgf4$^{+/+}$ and *Fgf4*$^{-/-}$ ESCs from the CD1 background (Wilder, P.J. 1997) were cultured in 2i/Lif (Ying et al. 2008) in accordance with established protocols (Mulas et al. 2019). Donor cells used in chimera experiments were labelled with dsRES via 1ug pPB-CAG-dsRED-pgk-Hyg (Guo et al. 2009), co-transfected with 2ug of transposase (pPBase; (Wang et al. 2008)), using Lipofectamine 2000. The cells were plated onto hygromycin resistant feeders and selected with 200 ug/ml hygromycin after 48 hours. Individual colonies were picked after 14 days based upon dsRed expression levels.

Embryo culture and chimera generation

Embryos were obtained from natural mating (C57BL/6xCBA). Detection of a copulation plug in the morning was used as confirmation of successful mating and indicated embryonic day (E) 0.5. Embryos were flushed from oviducts at E2.5 8-cell stage using M2 (Sigma-Aldrich, M7167).

Optimally, for chimera formation, embryos will be at the uncompacted 8-cell stage, enabling donor cells injected through the zona pellucida to become incorporated within the morula as it compacts. Embryo stages can vary within and between litters. Those recovered at the 4-cell stage can be cultured to the 8-cell stage. Embryos that have already compacted may be decompacted by brief culture in calcium-free medium for a few minutes before injection. Occasional abnormal embryos may be discarded, but otherwise all embryos are used. For the injection procedure, embryos and donor cells were placed in drops of M2 under oil on the microscope stage of the injection rig. Embryos were immobilised by means of a suction-mediated holding pipette. The desired number of separated ESCs were aspirated into the injection pipette. A small hole, just big enough to insert the injection pipette, was made in the zona pellucida opposite the holding pipette in a region of maximal space between blastomeres using a XYClone laser (Hamilton Thorne Biosciences). The injection pipette was gently pushed through the hole in the zona and the donor cells deposited away from the hole, between zona and host cells. Embryos were subsequently cultured for 48 hours in BlastAssist (Origio) as either a control or following ESC injection (Poueymirou et al. 2007). The experiments in **Fig. 2** and **Fig. 3** were conducted independently, on different days, and by different operators. Specifically, the injections in **Fig. 2** were performed by a more experienced operator, while those in **Fig. 3** were performed by a trainee. Reduced cell viability in **Fig. 3** may have contributed to the lower efficiency of chimera formation, likely due to the longer injection times associated with training. This could explain both the subtle shifts in lineage proportions and the observed reduction in PrE contribution in the 10+/+ group in **Fig. 3** that is not seen in **Fig. 2**.

This research has been regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body. Use of animals in this project was approved by the ethical review committee for the University of Cambridge, and relevant Home Office licences (Project licence number 80/2597 and number P76777883) were in place.

Immunohistochemistry

Embryos and chimeras were cultured to E2.5 + 48 hours post-harvest and fixed in 4% PFA in PBS for 15 minutes. They were rinsed in PBS with 3 mg/mL polyvinylpyrrolidone and blocked in 2% donkey serum, 0.01% BSA, 0.01% Tween20 in PBS for ~ 15 minutes. Primary antibodies were rat monoclonal anti-SOX2 (eBioscience, 14-9811-80) at 1:500 dilution, goat polyclonal anti-GATA4 (Santa Cruz Biotechnology, SC-1237) at 1:400 dilution, and DAPI at 1:10,000

dilution in blocking buffer. Embryos were incubated in primary antibodies at 4°C overnight, then rinsed 3 times in blocking buffer for at least 15 minutes each. Secondary antibodies conjugated with Alexa Fluor dyes of appropriate fluorophores and raised against the required hosts were diluted 1:500 in blocking buffer and embryos incubated for 1-2 hours at room temperature in the dark. They were subsequently rinsed as previously, equilibrated through increasing concentrations of mounting buffer, transferred to drops of concentrated mounting buffer on slides under coverslips subsequently sealed with nail varnish. Confocal images were acquired using an Andor Revolution XD spinning disk confocal microscope and Leica SP5.

Quantitative image analysis

Manual cell counting of ICMs from embryos and chimeras collected at E2.5+48 hrs was performed on confocal z-stacks using Fiji (ImageJ) with the 'Cell Counter' plugin. Individual nuclei were scored and assigned to lineages by eye, based on fluorescence marker expression. A nucleus was considered marker-positive if its signal was clearly above local background and showed nuclear enrichment for SOX2 and GATA4, or both nuclear and cytosolic enrichment for DsRed. SOX2-positive cells were classified as EPI, GATA4-positive cells as PrE, and cells double-positive for DsRed and SOX2 as donor-derived EPI. Semiautomated cell counting of whole embryos and chimeras collected at E2.5+48 hrs was performed with the MATLAB-based algorithm MINS (Lou et al. 2014) to perform 3D nuclear segmentation of the DAPI channel. Post analysis corrections for over- and under-segmentation was performed by manual segmentation in FIJI (Schindelin et al. 2012) using ROImanager to quantify (x,y,z) coordinates and fluorescence intensity for DAPI, SOX2, and GATA4. Each segment was sorted into SOX2 single positive, GATA4 single positive, or SOX2/GATA4 double negative populations by agglomerative hierarchical clustering using MATLAB (2016B) LINKAGE with WARD distances and the CLUSTER function. Individual segments were then clustered into 3D nuclei. Optimization of the clustering and number of nuclei was performed by Scree analysis and the elbow method, respectively. Nuclei that were segmented correctly by MINS were clustered into TE, PrE, and EPI in the same manner as the initial segment sorting (Strawbridge et al. 2023). Clustering was performed on a per-embryo basis.

Mathematical modelling

We modelled the mouse embryo cell population dynamics from the 8-cell stage (E2.5) to the late blastocyst stage (E4.5). Our mean-field model comprises a set of coupled ordinary differential equations (ODEs), which reflect our assumptions regarding cell state transitions,

proliferation and death (**Fig. 3A, B**). These equations govern the evolution in time (t) of the numbers of blastomeres (B), TE cells (T), unspecified ICM cells (C), PrE cells (P), EPI cells (E), and – for chimeric embryos (**Fig. 4A-C**) – the numbers of donor ESCs that are *Fgf4*$^{+/+}$ (D^+) or *Fgf4*$^{-/-}$ (D^-).

We assumed that all host cells proliferate with constant net per-capita rate α, which captures the net effect of division and death on cell number. This simplifying assumption is supported by a strong linear relationship between TE and ICM-derived cell numbers (unspecified ICM, PrE, and EPI), with a standardized effect size of $\beta = 1.1143$, suggesting that these lineages expand at comparable rates during this developmental window (**Fig. S1A**). We modelled the sorting of blastomeres to the embryo's interior, where they form the unspecified ICM, and exterior, where they epithelialize and become TE, as irreversible cell state transitions with constant per-capita rates $\rho\beta$ and $(1-\rho)\beta$, respectively. Here β denotes a constant overall per-capita rate of blastomere differentiation and ρ denotes the blastomere lineage bias towards the unspecified ICM. We modelled the specification of ICM into PrE and EPI as irreversible transitions. To account for the production by PrE of extracellular matrix proteins that help drive EPI specification, we assumed that the per-capita rate of transition from unspecified ICM to EPI, ζP^l, depends on the number of PrE cells present, with the parameter l allowing for non-linearity. Similarly, to account for the production by unspecified ICM and EPI of FGF4 that helps drive PrE specification, we assumed that the per-capita rate of transition from unspecified ICM to PrE, $\eta\,(C\,+\,E)^m$, depends on the numbers of unspecified ICM and EPI cells present, with the parameter m allowing for non-linearity. Here ζ captures the strength of feedback from PrE onto EPI specification, while η captures the strength of feedback from unspecified ICM and EPI onto PrE specification.

Accounting for the above processes leads to the following ODE system (**Fig. 1A**):

$$\frac{dB}{dt} = (\alpha - \beta)B, \tag{1}$$

$$\frac{dT}{dt} = \alpha T + (1-\rho)\beta B, \tag{2}$$

$$\frac{dC}{dt} = \left(\alpha - \zeta P^l - \eta\,(C\,+\,E)^m\right)C + \rho\beta B, \tag{3}$$

$$\frac{dE}{dt} = \alpha E + \zeta P^l C, \tag{4}$$

$$\frac{dP}{dt} = \alpha P + \eta\,(C\,+\,E)^m C. \tag{5}$$

Since our model describes the cell population dynamics starting from the 8-cell stage morula comprising only blastomeres, we impose the initial conditions $B(0) = 8, T(0) = C(0) = E(0) = P(0) = 0$. Our period of interest ends at $t = 48$ hours.

Chimera formation involves the injection of donor ESCs that are *Fgf4*$^{+/+}$ (D^+) or *Fgf4*$^{-/-}$ (D^-). This introduces additional contributions to the cell population dynamics. First, we assumed that donor cells proliferate at per-capita rate, α_D, which we allow to differ from that of host cells. Second, based on previous observations (Humięcka et al. 2016), we assumed that donor cells bias the rate of transition of blastomeres to unspecified ICM via crowding displacement, leading us to replace the parameter ρ with the decreasing function $\rho_0/(1 + a(D^+ + D^-))$. We write the sum $D^+ + D^-$ for convenience here; in practice, we only considered a model where either *Fgf4*$^{+/+}$ or *Fgf4*$^{-/-}$ ESCs are present. Third, to account for the production by FGF4+/+ ESCs of FGF4 that helps drive PrE specification, we modified the per-capita rate of transition from unspecified ICM to PrE from $\eta\ (C + E)$ to $\eta\ (C + E + D^+)$. Accounting for these additional processes led to the following ODE system describing cell population dynamics during chimera formation (**Fig. 1C**):

$$\frac{dB}{dt} = (\alpha - \beta)B, \tag{6}$$

$$\frac{dT}{dt} = \alpha T + \left(1 - \frac{\rho_0}{1 + a(D^+ + D^-)^n}\right)\beta B, \tag{7}$$

$$\frac{dC}{dt} = \left(\alpha - \zeta P^l - \eta\ (C\ +\ E\ +\ D_+)^m\right)C + \frac{\rho_0 \beta B}{1 + a(D^+ + D^-)^n}, \tag{8}$$

$$\frac{dE}{dt} = \alpha E + \zeta P^l C, \tag{9}$$

$$\frac{dP}{dt} = \alpha P + \eta(C\ +\ E\ +\ D^+)^m C, \tag{10}$$

$$\frac{dD^+}{dt} = \alpha_D D^+, \tag{11}$$

$$\frac{dD^-}{dt} = \alpha_D D^-. \tag{12}$$

We assumed that injection of donor ESCs occurs at the 8-cell stage morula, hence imposed the initial conditions $B(0) = 8$, $T(0) = C(0) = E(0) = P(0) = 0$, and either $D^+(0) = 10$, $D^-(0) = 0$ (injection of 10 FGF4+/+ ESCs), $D^+(0) = 15$, $D^-(0) = 0$ (injection of 15 *Fgf4*$^{+/+}$ ESCs), $D^+(0) = 0$, $D^-(0) = 10$ (injection of 10 *Fgf4*$^{-/-}$ ESCs), or $D^+(0) = 0$, $D^-(0) = 15$ (injection of 15 *Fgf4*$^{-/-}$ ESCs). Once again, our period of interest ends at $t = 48$ hours.

Equations (1)-(5) and (6)-(11) were solved numerically in MATLAB using an explicit Runge-Kutta method; see **Data Availability** for details on how to download our code.

Most of our model parameters cannot be directly measured and instead must be inferred from our data. We estimated parameters using Approximate Bayesian Computation (ABC) (Toni et al. 2009; Liepe et al. 2014), a likelihood-free method that iteratively compares a summary statistic from model simulations with given parameter values to the corresponding summary statistic from our data, and accepts those parameter values if these summary statistics are sufficiently close. By building up a set of accepted parameter values, ABC approximates their posterior distribution, allowing us to quantify our uncertainty in their values given our data. There are various well-established adaptations of ABC; we used a Markov Chain Monte Carlo approach (Marjoram et al. 2003). We used the summary statistic

$$S = \min_{t} \sum_{i} \quad \sum_{n} \quad (X_i(t) - X_i^n)^2 + \sum_{i} \quad \left(\max_{t} X_i(t) - \max_{n} X_i^n\right)^2 + B(t_{max}) + C(t_{max}),$$

where $X_i(t)$ denotes the value of the ith model variable at time t so that $X_1(t) = B(t)$, $X_2(t) = C(t)$ and so on; X_i^n denotes the value of the nth observation of the ith variable; and t_{max}. denotes the final time point in our model solution. The first term in S reflects our wish to have the model solution lie as close as possible to our data, accounting for the fact that time is implicit in our experimental observations (in other words, we are fitting the model in state space rather than as a time series). The second term in S reflects our wish to have the maximum value attained by each component of our model solution to be as close as possible to the corresponding maximum observation, as a way of helping to pin down timescales in our model. The third term in S reflects our wish to have as few as possible blastomeres and unspecified ICM cells present at the final time point in our model solution.

ACKNOWLEDGEMENTS

The authors thank Isaac Lundie-Fallon for support with collecting and imaging embryos, Sam Jameson and Polley Attesley for animal husbandry, and Néstor Saiz for discussions of complementary projects.

COMPETING INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

Conceptualization: S.E.S, A.G.F, J.N.; Data curation: S.E.S, A.K.S., P.H.; Formal analysis: S.E.S, A.K.S., P.H., A.G.F; Funding acquisition: S.E.S, A.G.F., J.N.; Investigation: S.E.S, A.K.S., A.G.F, J.N.; Methodology: S.E.S, A.G.F, J.N.; Project administration: S.E.S, A.G.F, J.N.; Resources: K.A.J., J.A., A.-K.H., J.N.; Software: S.E.S, A.K.S., P.H., A.G.F; Supervision: S.E.S, A.G.F, J.N.; Validation: S.E.S, A.K.S., J.A., A.G.F, J.N.; Visualization: S.E.S; Writing – original draft: S.E.S, A.G.F, J.N.; Writing – review & editing: S.E.S, A.G.F, J.N.

FUNDING

S.E.S. was supported by a Company of Biologists Travelling Fellowship (DEVTF-180513), a Sir Henry Wellcome Postdoctoral Fellowship (224070/Z/21/Z), and a University of Sheffield Strategic Research Fellowship in the Physics of Life and Quantitative Biology. A.G.F. was supported by the Biotechnology and Biological Sciences Research Council (BB/V018647/1 and BB/R016925/1). J.N. was supported by the University of Cambridge and the Biotechnology and Biological Sciences Research Council (BB/M023370/1).

DATA AVAILABILITY

MATLAB code for generating all results presented in **Figs. 1D-H**, **2C-G**, **3C-I**, **4D-F**, and **S1A-D**, **S2A-C**, and **S3A-C,** including all p-values, is freely available to download from GitHub (https://github.com/stanleystrawbridge/strawbridge_et_al_2024).

REFERENCES

Alexandrova, Stoyana, Tuzer Kalkan, Peter Humphreys, Andrew Riddell, Roberta Scognamiglio, Andreas Trumpp, and Jennifer Nichols. 2016. “Selection and Dynamics of Embryonic Stem Cell Integration into Early Mouse Embryos.” *Development (Cambridge, England)* 143 (1): 24–34.

Bessonnard, Sylvain, Laurane De Mot, Didier Gonze, Manon Barriol, Cynthia Dennis, Albert Goldbeter, Geneviève Dupont, and Claire Chazaud. 2014. “Gata6, Nanog and Erk Signaling Control Cell Fate in the Inner Cell Mass through a Tristable Regulatory Network.” *Development (Cambridge, England)* 141 (19): 3637–48.

Boroviak, Thorsten, Remco Loos, Paul Bertone, Austin Smith, and Jennifer Nichols. 2014. “The Ability of Inner-Cell-Mass Cells to Self-Renew as Embryonic Stem Cells Is Acquired Following Epiblast Specification.” *Nature Cell Biology* 16 (6): 516–28.

Boroviak, Thorsten, and Jennifer Nichols. 2014. “The Birth of Embryonic Pluripotency.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369 (1657). https://doi.org/10.1098/rstb.2013.0541.

Bradley, A., M. Evans, M. H. Kaufman, and E. Robertson. 1984. “Formation of Germ-Line Chimaeras from Embryo-Derived Teratocarcinoma Cell Lines.” *Nature* 309 (5965): 255–56.

Evans, M. J., and M. H. Kaufman. 1981. “Establishment in Culture of Pluripotential Cells from Mouse Embryos.” *Nature* 292 (5819): 154–56.

Feldman, B., W. Poueymirou, V. E. Papaioannou, T. M. DeChiara, and M. Goldfarb. 1995. “Requirement of FGF-4 for Postimplantation Mouse Development.” *Science* 267 (5195): 246–49.

Frankenberg, Stephen, François Gerbe, Sylvain Bessonnard, Corinne Belville, Pierre Pouchin, Olivier Bardot, and Claire Chazaud. 2011. “Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling.” *Developmental Cell* 21 (6): 1005–13.

Gardner, R. L., and J. Rossant. 1979. “Investigation of the Fate of 4·5 Day*post-Coitum*mouse Inner Cell Mass Cells by Blastocyst Injection.” *Development* 52 (1): 141–52.

Grabarek, Joanna B., Krystyna Zyzyńska, Néstor Saiz, Anna Piliszek, Stephen Frankenberg, Jennifer Nichols, Anna-Katerina Hadjantonakis, and Berenika Plusa. 2012. “Differential Plasticity of Epiblast and Primitive Endoderm Precursors within the ICM of the Early Mouse Embryo.” *Development (Cambridge, England)* 139 (1): 129–39.

Guo, Ge, Jian Yang, Jennifer Nichols, John Simon Hall, Isobel Eyres, William Mansfield, and Austin Smith. 2009. “Klf4 Reverts Developmentally Programmed Restriction of Ground State Pluripotency.” *Development (Cambridge, England)* 136 (7): 1063–69.

Hillman, N., M. I. Sherman, and C. Graham. 1972. “The Effect of Spatial Arrangement on Cell Determination during Mouse Development.” *Journal of Embryology and Experimental Morphology* 28 (2): 263–78.

Humięcka, Monika, Magdalena Krupa, Maria M. Guzewska, Marek Maleszewski, and Aneta Suwińska. 2016. “ESCs Injected into the 8-Cell Stage Mouse Embryo Modify Pattern of Cleavage and Cell Lineage Specification.” *Mechanisms of Development* 141 (August):40–50.

Kang, Minjung, Vidur Garg, and Anna-Katerina Hadjantonakis. 2017. “Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2.” *Developmental Cell* 41 (5): 496–510.e5.

Kang, Minjung, Anna Piliszek, Jérôme Artus, and Anna-Katerina Hadjantonakis. 2013. “FGF4 Is Required for Lineage Restriction and Salt-and-Pepper Distribution of Primitive Endoderm Factors but Not Their Initial Expression in the Mouse.” *Development* 140 (2): 267–79.

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding Groups in Data*. Edited by Leonard Kaufman and Peter J. Rousseeuw. 99th ed. Probability & Mathematical Statistics S. Nashville, TN: John Wiley & Sons.

Liepe, Juliane, Paul Kirk, Sarah Filippi, Tina Toni, Chris P. Barnes, and Michael P. H. Stumpf. 2014. “A Framework for Parameter Estimation and Model Selection from Experimental Data in Systems Biology Using Approximate Bayesian Computation.” *Nature Protocols* 9 (2): 439–56.

Lou, Xinghua, Minjung Kang, Panagiotis Xenopoulos, Silvia Muñoz-Descalzo, and Anna-Katerina Hadjantonakis. 2014. “A Rapid and Efficient 2D/3D Nuclear Segmentation Method for Analysis of Early Mouse Embryo and Stem Cell Image Data.” *Stem Cell Reports* 2 (3): 382–97.

Maître, Jean-Léon, Hervé Turlier, Rukshala Illukkumbura, Björn Eismann, Ritsuya Niwayama, François Nédélec, and Takashi Hiiragi. 2016. “Asymmetric Division of Contractile Domains Couples Cell Positioning and Fate Specification.” *Nature* 536 (7616): 344–48.

Marjoram, Paul, John Molitor, Vincent Plagnol, and Simon Tavare. 2003. “Markov Chain Monte Carlo without Likelihoods.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (26): 15324–28.

Martin, G. R. 1981. “Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 78 (12): 7634–38.

Mulas, Carla, Tüzer Kalkan, Ferdinand von Meyenn, Harry G. Leitch, Jennifer Nichols, and Austin Smith. 2019. “Defined Conditions for Propagation and Manipulation of Mouse Embryonic Stem Cells.” *Development (Cambridge, England)* 146 (6). https://doi.org/10.1242/dev.173146.

Nichols, Jennifer, Jose Silva, Mila Roode, and Austin Smith. 2009. “Suppression of Erk Signalling Promotes Ground State Pluripotency in the Mouse Embryo.” *Development* 136 (19): 3215–22.

Nissen, Silas Boye, Marta Perera, Javier Martin Gonzalez, Sophie M. Morgani, Mogens H. Jensen, Kim Sneppen, Joshua M. Brickman, and Ala Trusina. 2017. “Four Simple Rules That Are Sufficient to Generate the Mammalian Blastocyst.” *PLoS Biology* 15 (7): e2000737.

Plusa, Berenika, Anna Piliszek, Stephen Frankenberg, Jérôme Artus, and Anna-Katerina Hadjantonakis. 2008. “Distinct Sequential Cell Behaviours Direct Primitive Endoderm Formation in the Mouse Blastocyst.” *Development* 135 (18): 3081–91.

Poueymirou, William T., Wojtek Auerbach, David Frendewey, Joseph F. Hickey, Jennifer M. Escaravage, Lakeisha Esau, Anthony T. Doré, et al. 2007. “F0 Generation Mice Fully Derived from Gene-Targeted Embryonic Stem Cells Allowing Immediate Phenotypic Analyses.” *Nature Biotechnology* 25 (1): 91–99.

Saiz, Néstor, Laura Mora-Bitria, Shahadat Rahman, Hannah George, Jeremy P. Herder, Jordi Garcia-Ojalvo, and Anna-Katerina Hadjantonakis. 2020. “Growth-Factor-Mediated Coupling between Lineage Size and Cell Fate Choice Underlies Robustness of Mammalian Development.” *eLife* 9 (July). https://doi.org/10.7554/eLife.56079.

Saiz, Néstor, Kiah M. Williams, Venkatraman E. Seshan, and Anna-Katerina Hadjantonakis. 2016. “Asynchronous Fate Decisions by Single Cells Collectively Ensure Consistent Lineage Composition in the Mouse Blastocyst.” *Nature Communications* 7 (November):13463.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. “Fiji: An Open-Source Platform for Biological-Image Analysis.” *Nature Methods* 9 (7): 676–82.

Strawbridge, Stanley E., Agata Kurowski, Elena Corujo-Simon, Alastair N. Fletcher, Jennifer Nichols, and Alexander G. Fletcher. 2023. “insideOutside: An Accessible Algorithm for Classifying Interior and Exterior Points, with Applications in Embryology.” *Biology Open* 12 (9). https://doi.org/10.1242/bio.060055.

Tarkowski, A. K., and J. Wróblewska. 1967. “Development of Blastomeres of Mouse Eggs Isolated at the 4- and 8-Cell Stage.” *Journal of Embryology and Experimental Morphology* 18 (1): 155–80.

Toni, Tina, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P. H. Stumpf. 2009. “Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems.” *Journal of the Royal Society, Interface / the Royal Society* 6 (31): 187–202.

Wang, Wei, Chengyi Lin, Dong Lu, Zeming Ning, Tony Cox, David Melvin, Xiaozhong Wang, Allan Bradley, and Pentao Liu. 2008. “Chromosomal Transposition of PiggyBac in Mouse Embryonic Stem Cells.” *Proceedings of the National Academy of Sciences* 105 (27): 9290–95.

Wilder, P.J. 1997. “Inactivation of the FGF-4 Gene in Embryonic Stem Cells Alters the Growth And/or the Survival of Their Early Differentiated Progeny.” *Developmental Biology* 192 (2): 614–29.

Yamanaka, Yojiro, Fredrik Lanner, and Janet Rossant. 2010. “FGF Signal-Dependent Segregation of Primitive Endoderm and Epiblast in the Mouse Blastocyst.” *Development* 137 (5): 715–24.

Yanagida, Ayaka, Elena Corujo-Simon, Christopher K. Revell, Preeti Sahu, Giuliano G. Stirparo, Irene M. Aspalter, Alex K. Winkel, et al. 2022. “Cell Surface Fluctuations Regulate Early Embryonic Lineage Sorting.” *Cell* 185 (5): 777–93.e20.

Ying, Qi-Long, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. 2008. “The Ground State of Embryonic Stem Cell Self-Renewal.” *Nature* 453 (7194): 519–23.

Figures

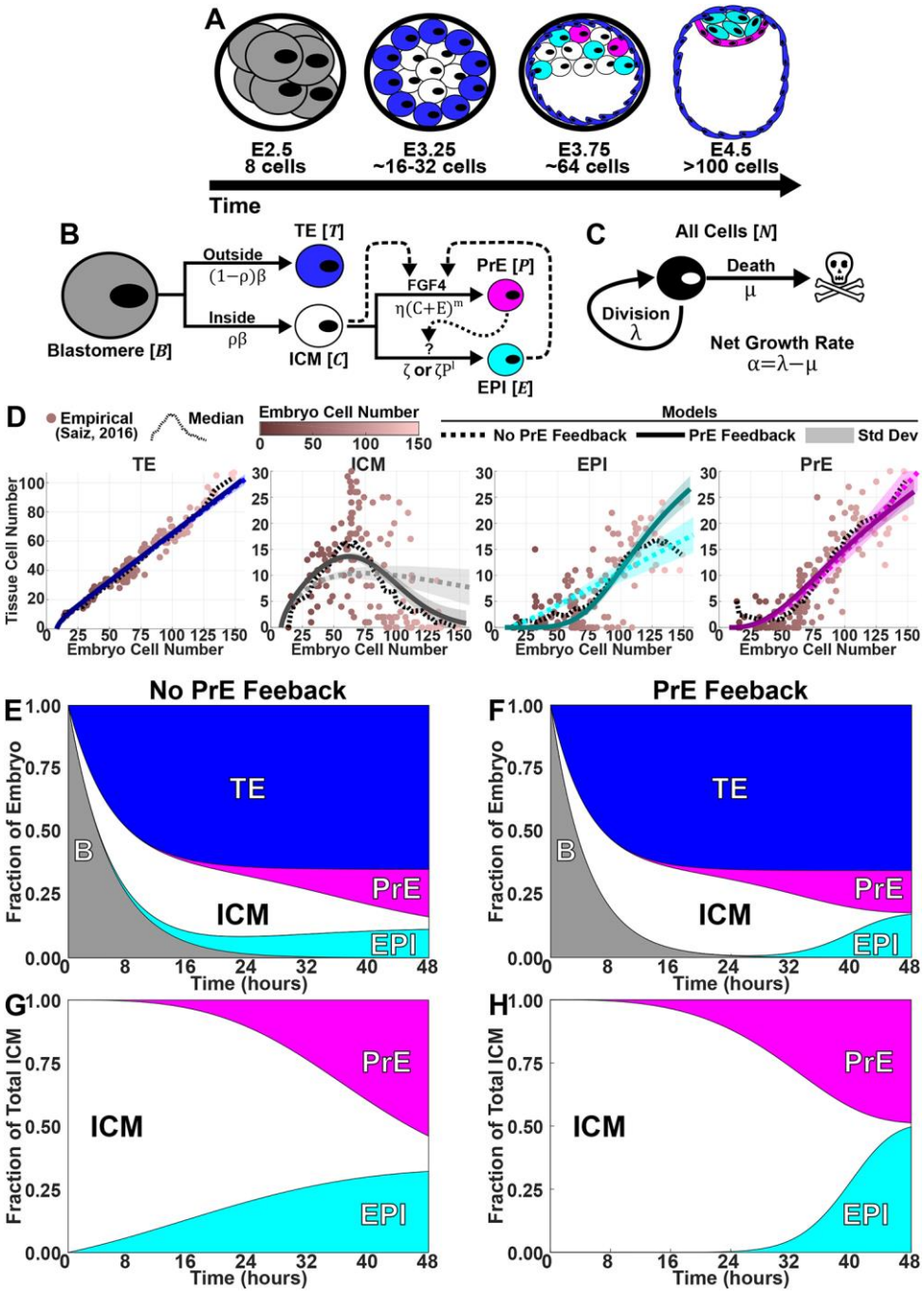

Fig. 1. Quantitative modelling implicates PrE feedback in ICM specification. A. Preimplantation mouse development from E2.5 (8-cell morula) to E4.5 (late blastocyst). **B.** Model of cell-state transitions during the first two fate decisions in mouse

development. Solid lines: transitions; dashed line: FGF4 feedback; dotted line: proposed PrE feedback. **C.** All cells grow at net rate α. **D.** Numerical solutions using median posterior parameters from models without (dotted) and with (solid) PrE feedback, overlaid on data (dots) from (Saiz et al. 2016) and binned median. **E-H.** Reconstructed time course of relative tissue sizes in whole embryo (**E, F**) and ICM (**G, H**) for models without (**E, G**) and with (**F, H**) PrE feedback.

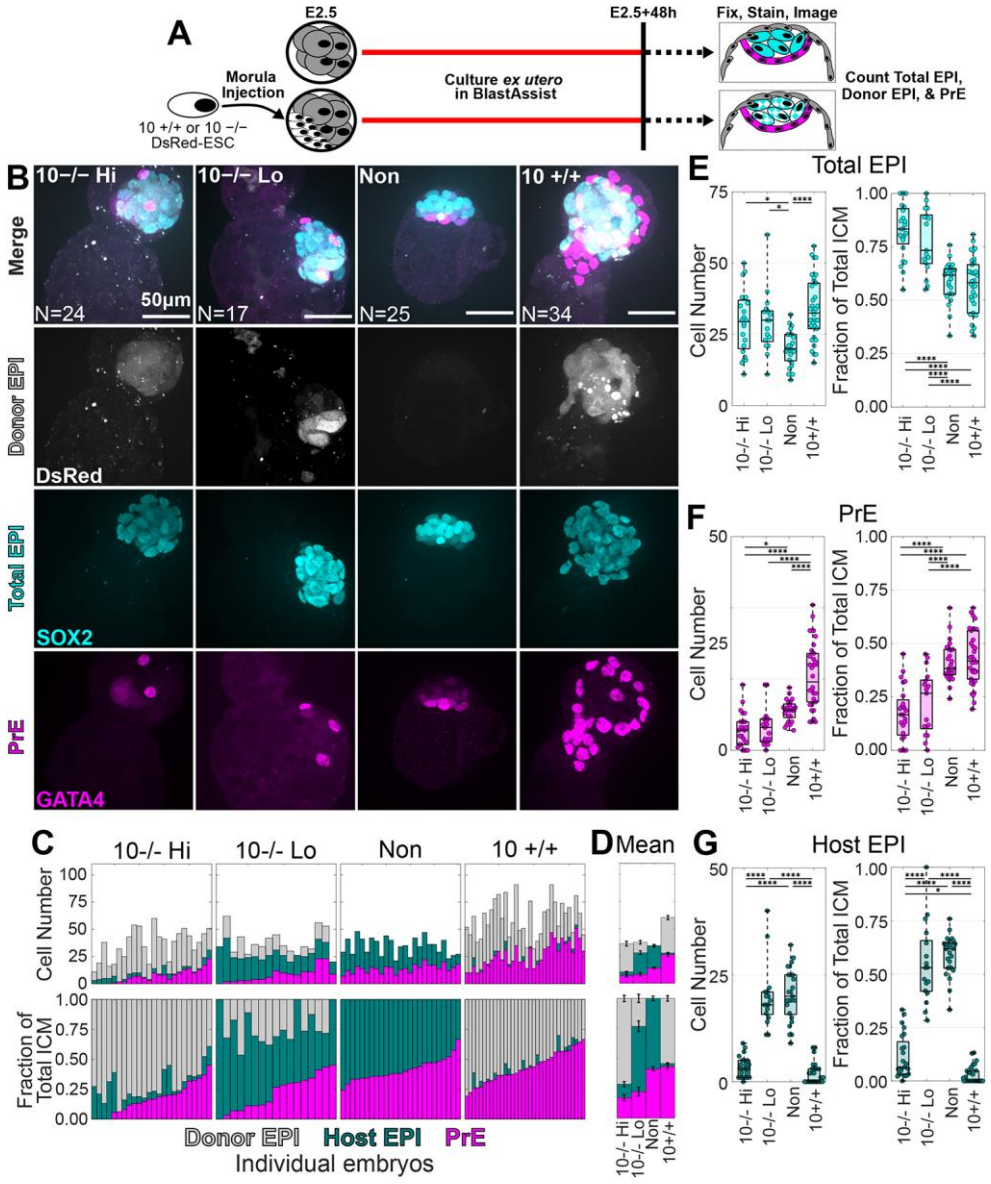

Fig. 2. Both *Fgf4*$^{+/+}$ and *Fgf4*$^{-/-}$ donor ESCs can impede host contribution to the EPI. A. Schematic of ICM imaging after injection of 10 donor cells. **B.** Representative maximum intensity projections of confocal z-stacks for chimeras injected with 10 *Fgf4*$^{-/-}$ ESCs showing high (Hi, N=24) or low (Lo, N=17) donor contribution, non-injected embryos (N=25), and 10 *Fgf4*$^{+/+}$ ESC-injected chimeras (N=34). **C, D.** Stacked bar plots showing individual (**C**) and mean ± SEM (**D**) ICM composition by cell number (top) and fraction (bottom). Magenta: GATA4+ PrE; dark green: SOX2+ host EPI; gray: DsRed+/SOX2+ donor cell-derived EPI. Bars sorted by PrE fraction. **E-G.** Box/swarm

plots showing cell number (left) and ICM fraction (right) for total EPI (**E**), PrE (**F**), and host=derived EPI (**G**). Pairwise comparisons by N-way ANOVA, p-values: *, $0.05 \geq p > 0.01$; **, $0.01 \geq p > 0.001$; ***, $0.001 \geq p > 0.0001$; ****, $0.0001 \geq p$. See **Table S1** for full summary statistics and exact p-values.

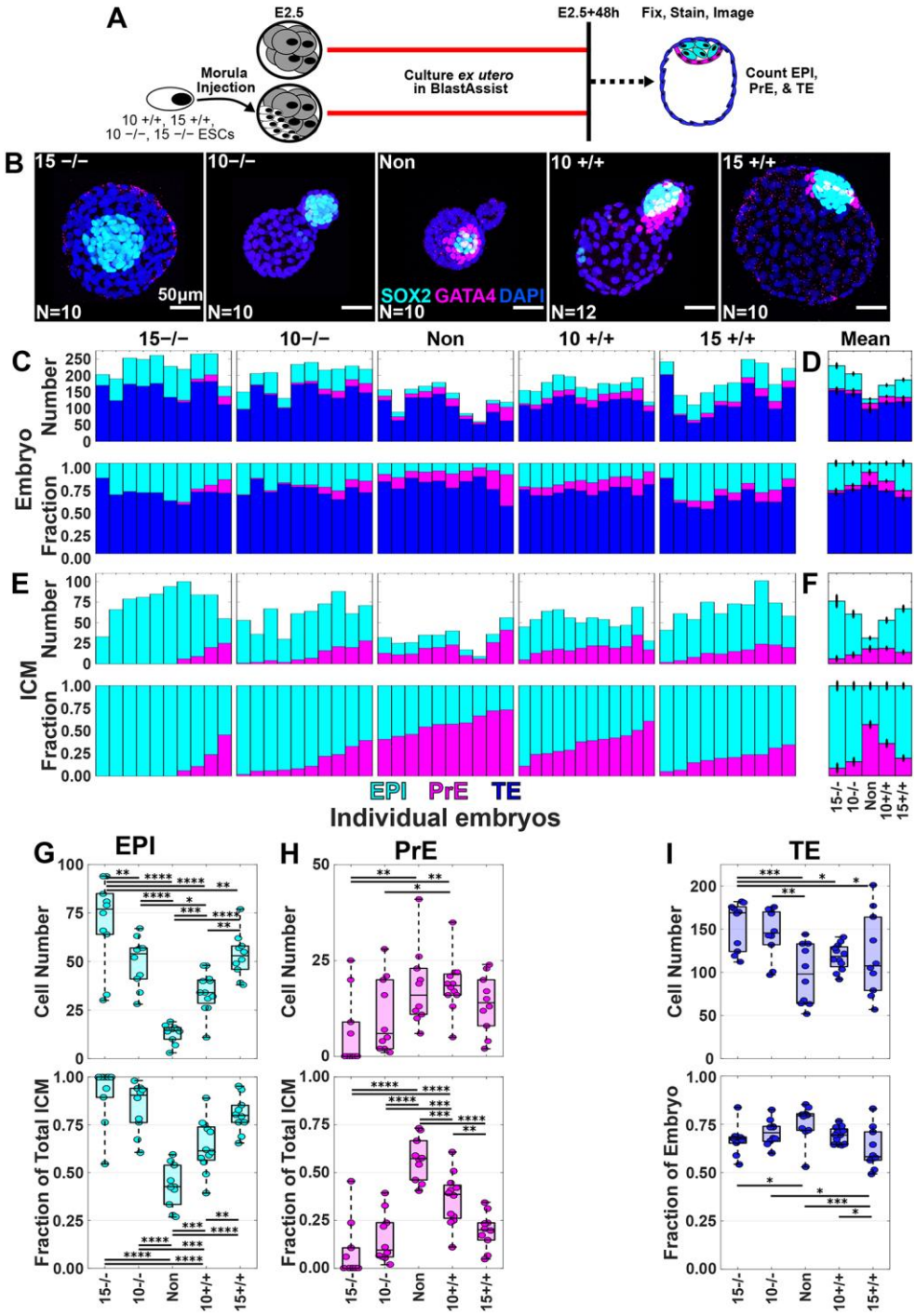

Fig. 3. Embryos injected with *Fgf4*$^{-/-}$ donor ESCs have fewer PrE cells and more TE cells. A. Schematic of whole-embryo imaging after injection of 10 or 15 donor cells. **B.** Representative maximum intensity projections of confocal z-stacks for non-injected embryos (N=10), 10 and 15 *Fgf4*$^{-/-}$ ESC-injected chimeras (N=10,10), and 10 and 15

Fgf4$^{+/+}$ ESC-injected chimeras (N=12,10). **C, D.** Stacked bar plots of individual (**C**) and mean (±SEM) (**D**) ICM composition for chimeras and embryos showing cell number (top) and whole-embryo fraction (bottom). Magenta: GATA4+ PrE; cyan: SOX2+ EPI; blue: double-negative TE. Bars sorted by PrE fraction. **E,F.** Stacked bar plots for individual (**E**) and mean ±SEM (**F**) ICM composition for chimeras and embryos showing cell number (top) and fraction of the ICM (bottom). **G,H.** Box/swarm plots showing cell number (top) and ICM fraction (bottom) for EPI (**E**), PrE (**F**), and host-derived EPI (**G**). **I.** Box / swarm plots showing cell number (top) and whole-embryo fraction (bottom) for TE. Pairwise comparisons by N-way ANOVA, p-values: *, $0.05 \geq p > 0.01$; **, $0.01 \geq p > 0.001$; ***, $0.001 \geq p > 0.0001$; ****, $0.0001 \geq p$. See **Table S1** for full summary statistics and exact p-values.

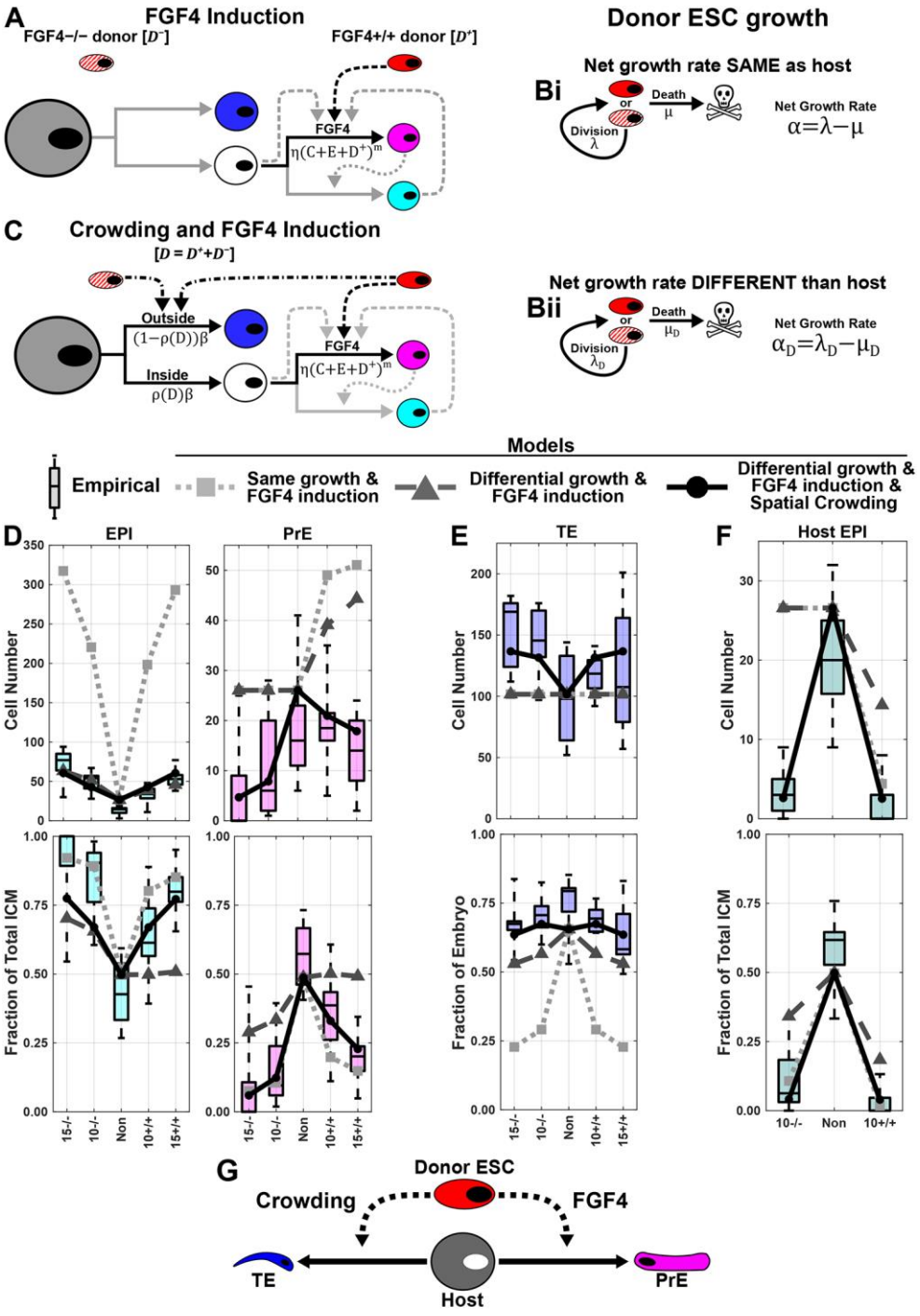

Fig. 4. Quantitative modelling suggests spatial crowding plays a role in host EPI exclusion during chimera formation. A. FGF4 induction model for chimera formation. Gray: blastomere; blue: TE; white: unspecified ICM; cyan: EPI; magenta: PrE; red: *Fgf4*$^{+/+}$ donor ESC; red/white: *Fgf4*$^{-/-}$ donor ESC; solid lines: cell-state transitions; dashed lines: FGF4 feedback; dotted line: posited PrE feedback; dash-dotted line:

spatial crowding feedback. **B.** Donor ESC net growth rate: (i) fixed at host cell rate; (ii) inferred. **C.** FGF4 induction and spatial crowding model for chimera formation. **D, E.** Simulations (lines and markers) superimposed on data from **Fig. 3G-I** used to infer additional parameters. **F.** Simulations superimposed on host EPI data from **Fig. 2G**. **G.** Proposed host EPI exclusion model.

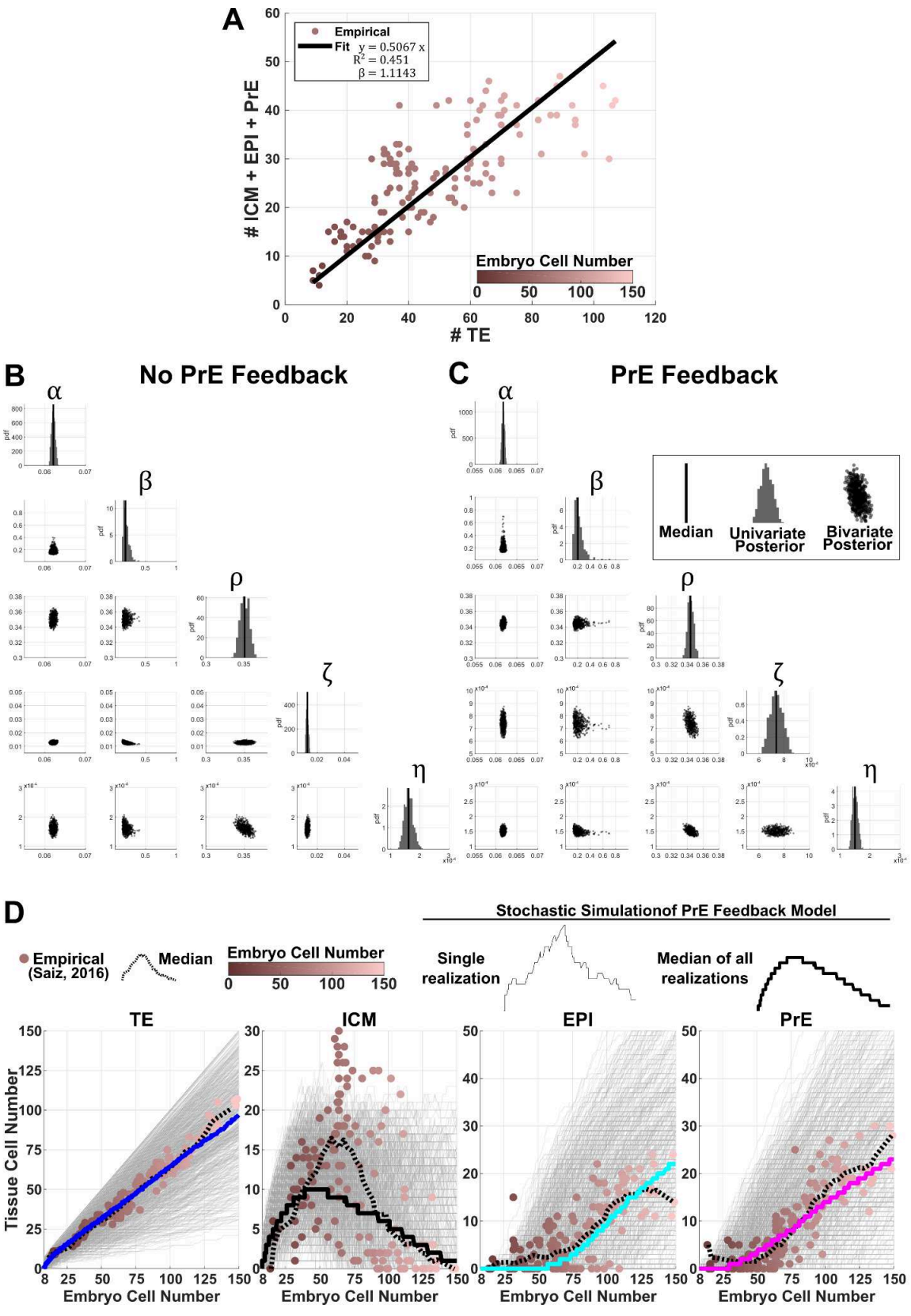

Fig. S1. Posterior parameter distributions and stochastic simulations for embryogenesis. A. Scatter plot showing the relationship between the number of TE and number of cells in the ICM compartment (unspecified ICM + EPI + PrE) per embryo data from (Saiz et al. 2016). A linear model was fit without intercept yielding $y = 0.5067x$, with $R^2 = 0.4510$, p-value $= 2.436 \times 10^{-80}$, and Standardized effect size $\beta = 1.1143$. Each point is colored by total embryo cell number. **B,C.** Univariate (on diagonal histograms, vertical line shows median) and bivariate (off diagonal scatter plots) posterior distributions of inferred parameters for models without (**B**) and with (**C**) PrE feedback on ICM specification. Parameters are: $\boldsymbol{\alpha}$, net growth rate; $\boldsymbol{\beta}$, blastomere specification rate; $\boldsymbol{\rho}$ blastomere bias towards the unspecified ICM fate; $\boldsymbol{\zeta}$, coefficient for ICM to EPI specification; and $\boldsymbol{\eta}$, coefficient for ICM to PrE specification. Not shown are posteriors for feedback parameters l, m, which both take on the value of 2 for all accepted parameter sets. Priors were uniform over the ranges shown in the plots. ABC thresholds (ε) were 17,300 for the model without PrE feedback and 13,500 for the model with PrE feedback. **D.** Stochastic simulations of the embryo model using the Gillespie algorithm, run with the posterior median parameters from the PrE feedback model. Shown are 1,000 independent realizations (light gray line) and median of all realizations (think colored line), overlaid on empirical data (dots) from (Saiz et al. 2016) and the median (50 bins).

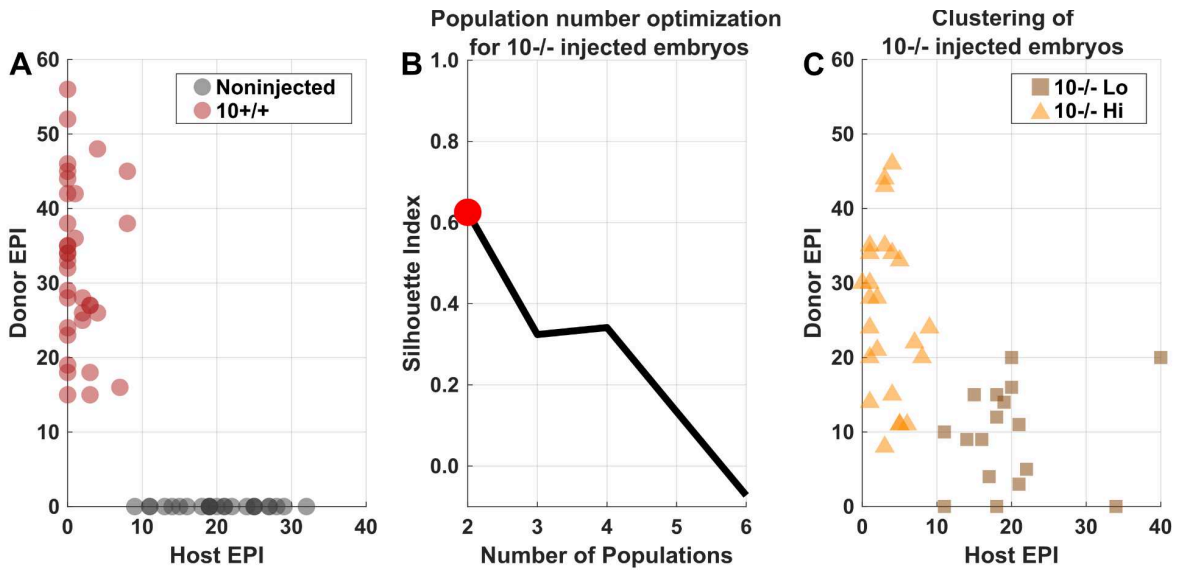

Fig. S2. k-means clustering identifies two populations in the 10 *Fgf4*$^{-/-}$ donor ESC injected embryos. A. EPI composition of noninjected (black circles) and 10 *Fgf4*$^{+/+}$ donor ESC injected (red circles) embryos. **B.** Black line: silhouette index from k-means clustering for different numbers of populations of *Fgf4*$^{+/+}$ donor ESC injected embryos; red dot: optimal number of populations. **C.** EPI composition of *Fgf4*$^{-/-}$ donor ESC injected embryos grouped by k-means clustering with optimal number of populations revealing a low (Lo) (brown squares) and high (Hi) (yellow triangles) donor cell contribution population.

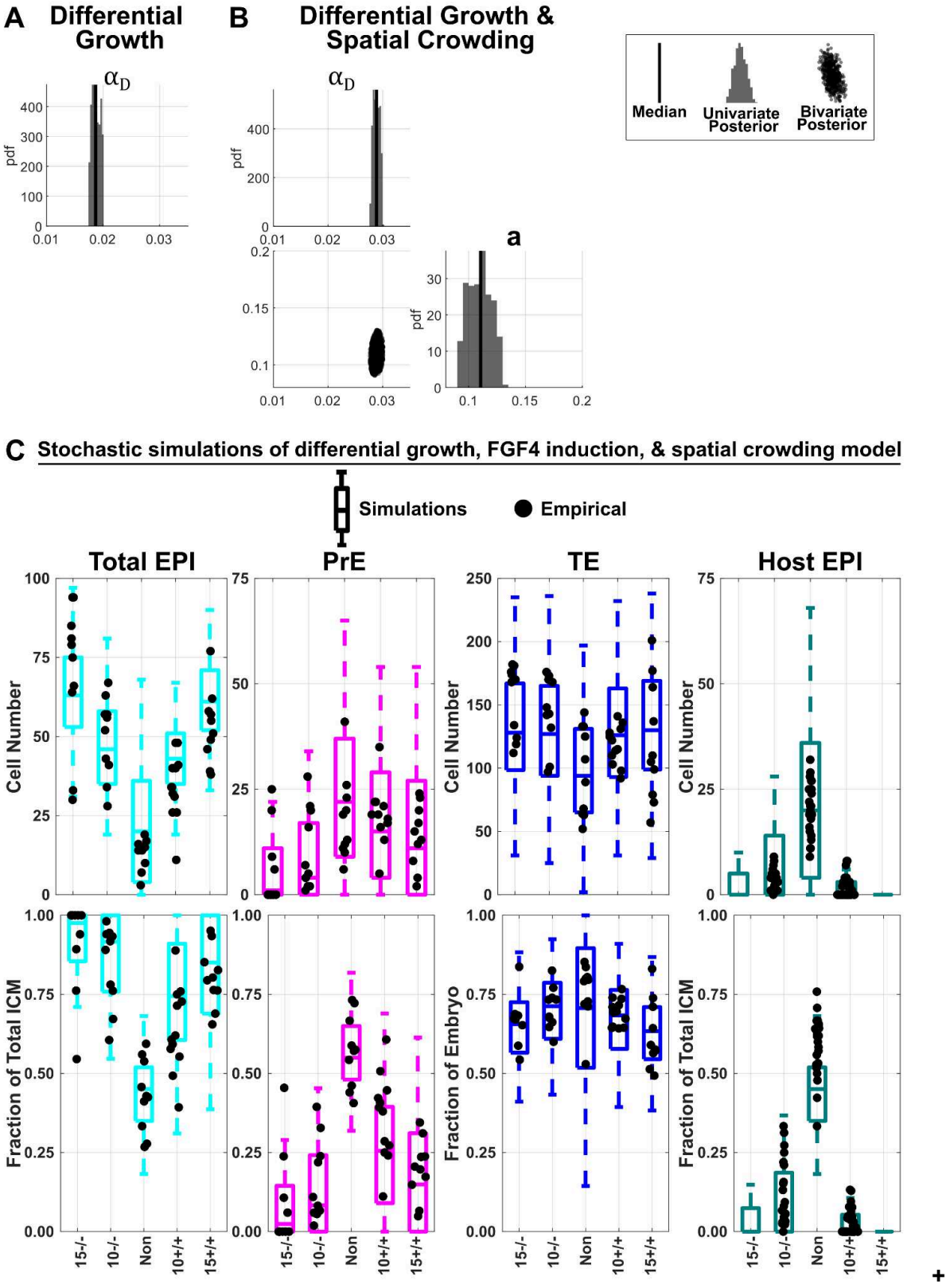

Fig. S3. Posterior distributions of inferred parameters for different models of chimera formation. Univariate (on diagonal histograms, vertical line shows median) and bivariate (off diagonal scatter plots) posterior distributions of inferred parameters for models with differential growth for donor cells (**A**) and with differential growth and spatial crowding (**B**). Parameters are: α_D, donor cell net growth rate; **a**, crowding factor. Not shown is the posterior for feedback parameter n which took on the value of 1 for all accepted parameter sets in the differential growth and spatial crowding model. Priors were uniform over the ranges shown in the plots. ABC thresholds (ε) were 125,000 for the model with differential growth and 62,000 for the model with differential growth and spatial crowding. **C.** Stochastic simulations of the chimera model using the Gillespie algorithm, run with the posterior median parameters from the differential growth, FGF4 induction, and spatial crowding model. Shown are 1,000 independent realizations (box plots).

Table S1. Summary of statistical analyses corresponding to Figures 2E–G and 3G–I. Summary statistics (mean, standard deviation, and N) and p-values from N-way ANOVA with Tukey's honestly significant difference (HSD) post hoc test are provided for each experimental condition. Statistical significance was assessed at α = 0.05.

Available for download at
https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.204518#supplementary-data