bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Generative model for the first cell fate bifurcation in mammalian development

Maria Avdeeva*[1,4], Madeleine Chalifoux*[2,3], Bradley Joyce[3], Stanislav Y. Shvartsman[1,2,3], and Eszter Posfai[3,4]

[1] Center for Computational Biology, Flatiron Institute, Simons Foundation, New York, New York, USA
[2] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, USA
[3] Department of Molecular Biology, Princeton, New Jersey, USA
[4] Correspondence: mavdeeva@flatironinstitute.org, eposfai@princeton.edu

Abstract. The first cell fate bifurcation in mammalian development directs cells toward either the trophectoderm (TE) or inner cell mass (ICM) compartments in preimplantation embryos. This decision is regulated by the subcellular localization of a transcriptional co-activator YAP and takes place over several progressively asynchronous cleavage divisions. As a result of this asynchrony and variable arrangement of blastomeres, reconstructing the dynamics of the TE/ICM cell specification from fixed embryos is extremely challenging. To address this, we developed a live imaging approach and applied it to measure pairwise dynamics of nuclear YAP and its direct target genes, CDX2 and SOX2, key transcription factors of TE and ICM, respectively. Using these datasets, we constructed a generative model of the first cell fate bifurcation, which reveals the time-dependent statistics of the TE and ICM cell allocation. In addition to making testable predictions for the joint dynamics of the full YAP/CDX2/SOX2 motif, the model revealed the stochastic nature of the induction timing of the key cell fate determinants and identified the features of YAP dynamics that are necessary or sufficient for this induction. Notably, temporal heterogeneity was particularly prominent for SOX2 expression among ICM cells. As heterogeneities within the ICM have been linked to the initiation of the second cell fate decision in the embryo, understanding the origins of this variability is of key significance. The presented approach reveals the dynamics of the first cell fate choice and lays the groundwork for dissecting the next cell fate bifurcations in mouse development.

1 Introduction

Early development of most well-studied model organisms follows a largely deterministic developmental pattern where differentiation can be attributed to pre-existing asymmetries derived from sperm entry and/or nonuniformly distributed maternal factors [1]. In the mouse, the leading model for studying mammalian development, such early asymmetries are not major factors contributing to differentiation [2,3]. Instead, differentiation, first detected after the fourth cleavage that generates the 16-cell morula, is driven by self-organization [4]. Because mouse embryos do not follow an invariant cleavage pattern, the morulae vary in shapes and arrangements of blastomeres [5,6,7,8,9]. This geometric variability has been shown to be an essential component of differentiation, with geometric cues driving cells towards their respective fates [10,11,12]. Furthermore, spatial and temporal variability in cleavage patterns

* equal contribution

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and resulting gene expression variability have been proposed to be a driver of robustness in the first cell fate decisions [9,8,13,14]. As a consequence of this unique mode of development, quantitative models of the early mouse embryo must be distinct from the essentially deterministic models that suffice for embryos that depend on the invariant pattern of cleavages and preexisting asymmetries.

Here we establish such a model for the first cell fate bifurcation in the mouse embryo, as cells are directed to either the inner cell mass (ICM), precursor of the fetus and certain extraembryonic tissues, or the trophectoderm (TE), precursor of the placental cell types (Fig. 1A). The specification of these two cell types is associated with the expression of well-known lineage-specific transcription factors (TFs), such as SOX2 (ICM) and CDX2 (TE) [15,16]. Both CDX2 and SOX2 are regulated by the Hippo signaling pathway and its downstream effector, the Yes-associated protein (YAP), where YAP serves as an activator of *Cdx2* and a repressor of *Sox2* expression [17,15,18,19] (Fig. 1B). YAP activity is determined by its subcellular localization, which in turn is regulated by apicobasal polarity of cells, with polarized cells localizing YAP to the nucleus and apolar cells sequestering YAP in the cytoplasm [17,20,21].

Segregation of ICM and TE fates largely takes place between the 8- to 32-cell stages. At the 8-cell stage, gene expression variability among cells is low, and individual cells display bipotency towards ICM and TE fates [22]. By the 32-cell stage, TE cells line the outside of the embryo and ICM cells are located on the inside. This sequence of events was largely deduced from images of fixed embryos, which offer only snapshots of nuclear YAP and its transcriptional effects, with limited live imaging datasets of individual components available to date [23,24,25]. Using these datasets for quantitative understanding and modeling of the underlying dynamics is a highly challenging task [26,13,27,28,29,30]. Using live imaging together with computational cell segmentation and tracking, we acquire a view of the simultaneous dynamics of nuclear YAP and its downstream targets within lineages of all blastomeres. To model variability in the observed dynamics, we choose a dynamic Bayesian network approach which is a well-established strategy for modeling gene expression time series [31,32,33,34,35]. The resulting generative model reveals the dynamics and variability of the first cell fate choice in mammalian development.

2 RESULTS

Live imaging YAP and its target genes

Taking advantage of the simplicity of the network that controls the first mammalian cell fate choice, as well as the small cell number at which this choice takes place, we sought to characterize the dynamics of YAP and its targets by live imaging. As the first step towards this goal, we established or used already existing fluorescent reporters of YAP (YAP-miRFP670; [25]), CDX2 (CDX2-eGFP; [23]) and SOX2 (mScarlet-I-SOX2; this study). To generate a SOX2 reporter mouse line, we targeted mScarlet-I to the N-terminus of the endogenous *Sox2* allele (Supp. Fig. S1A). mScarlet-I expression displayed the expected pattern of endogenous SOX2 based on previous analysis of fixed samples, with expression initiating in a few cells at the late 16-cell stage and mScarlet-I present in all ICM cells at the blastocyst stage (Fig. 1C, Supp. Fig. S1E,F). Furthermore, treatment of embryos with the Rho-associated kinase (ROCK) inhibitor caused an increase in the number of cells expressing mScarlet-I-SOX2, in agreement with a previous report ([18], Supp. Fig. S1B,C,D). These

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

data indicate that the established reporter is suitable for monitoring endogenous SOX2 dynamics.

We used light sheet microscropy to image the joint dynamics of YAP-miRFP670 and CDX2-eGFP or YAP-miRFP670 and mScarlet-I-SOX2, along with H2B-miRFP720 to identify cell nuclei (Fig. 1C,D). Nuclei were then segmented and tracked using our previously established pipeline [36] to construct complete lineage trees of developing embryos. To quantify the dynamics of YAP and its targets within each lineage, fluorescent intensities for each reporter were extracted from nuclear masks (Fig. 1E,F, Methods). Every embryo in our dataset was imaged from the 8- to the end of the 32-cell stage, with the final cells defining 32 branches of expression histories. For each tree branch of every embryo, we sought to ultimately characterize the dynamics of YAP and its targets in the (y, S, C) phase space, where y is the nuclear concentration of YAP, and S and C are the total nuclear levels of SOX2 and CDX2, respectively.

Cells in mouse embryos exhibit random, normally distributed timing of cleavages at the 16- and 32-cell stages [9]. We normalized the time for each branch of every embryo by warping it to align cell division times at every cell cycle (Methods). The expression of both YAP targets was heterogeneous, both in the timing and levels of expression (Fig. 2A,B). This was accompanied by heterogeneity in the corresponding YAP trajectories.

As the first step towards modeling the observed dynamics, we coarse-grained the phase space by distinguishing only two levels for each variable, which effectively views YAP as either cytoplasmic or nuclear and each of its targets as either expressed or not. To this end, we first separated 16- and 32-cell stages into early (E) and late (L) stages resulting in 5 stages of interest: 8, 16E, 16L, 32E, 32L, indexed by $i = 0 \dots 4$. For every stage, the expression levels for YAP, SOX2 and CDX2 were binarized (Methods). In particular, YAP concentration was z-scored at every timepoint and averaged over each stage; pooling the data, we defined universal stage-dependent thresholds (Supp. Fig. S2B) to characterize YAP as either nuclear (1) or cytoplasmic (0). To binarize the expression of each downstream TF, after appropriate normalization, we applied a stage-independent threshold to its values at the end of 16L, 32E, and 32L to define positive (1) and negative (0) cells at these stages (N=6 embryos for SOX2, N=9 embryos for CDX2, Supp. Fig. S2C,D). We characterized YAP as nuclear for all cells at stage 8 and all cells as negative for both TFs at stages 8 and 16E.

As a result of the above coarse-graining, imaging data from every embryo was decomposed into 32 binarized trajectories indexed by stage, $i = 0 \dots 4$. More precisely, YAP dynamics and expression of its downstream gene g were each described by binary sequences of length 5, $y = \{y_i\}_{i=0\dots4} \in \{0,1\}^5$, and $G = \{G_i\}_{i=0\dots4} \in \{0,1\}^5$ for $G = S, C$ (Fig. 2A,B). Coarse-graining allowed us to classify the trajectories by their time of expression induction, i.e., $\min\{i : G_i = 1\}$. For $G = S, C$, we denoted the classes of trajectories inducing expression at 16L, 32E and 32L stages as $G^+_{16L}, G^+_{32E}, G^+_{32L}$, respectively. Note that, after induction, a trajectory can potentially switch back to 0. Such loss of expression was very rare though possible for both TFs (see, e.g., Fig. 2A). The rest of the trajectories never induced the TF, and we denoted the corresponding class as G^-. We observed that trajectory classes might exhibit different YAP coarse-grained dynamics (Fig. 2A,B). To systematically explore the role of YAP in setting the times at which cells express SOX2 and CDX2, we proceeded to model the coarse-grained dynamics in (y, G) phase space using dynamic Bayesian networks.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Bayesian modeling of pairwise dynamics

Bayesian networks model the data using directed acyclic graphs that define factorization of the joint probability distribution of the nodes providing, for every node, the probability distribution conditional on its parents [33]. In our case, each node corresponds to the levels of nuclear YAP, y_i, or one of its targets, G_i, at a timepoint i. For each node, we chose its parents from the nodes corresponding to the same or the previous timepoint. Thus, we adopted a dynamic Bayesian network framework, with $\{y_i\}_{i=0\ldots4}$ and $\{G_i\}_{i=0\ldots4}$ in the nodes.

Having binarized all the variables, we assumed Bernoulli distribution for every node. We assumed that YAP localization is memory-less and therefore y can be modeled as a Markov chain. YAP localization affects transcription of downstream genes. Assuming no prior information on the timescales of molecular processes that connect subcellular YAP localization and the expression of its targets, we explored several architectures for the Bayesian network downstream of y (Fig. 3A). The networks were assumed to be non-homogeneous, with time-dependent conditional probability distributions (CPDs). The simplest network assumed fast effect of YAP localization on the downstream protein concentration, captured by $\{(y_i, G_i)\}_{i=0\ldots4}$ edges, and conditional (on y) independence of G_{i+1} on G_i for every i (M1 model). We also considered networks that connect G nodes into another chain. These models include fast (M2), or slow, captured by $\{(y_i, G_{i+1})\}_{i=0\ldots3}$ edges, timescales for the effect of y on G (M3). Finally, we included a network that combined the fast and slow timescales (M4). To summarize, we chose 4 candidate models to test for every dataset.

For each (y, G) pair, we sought to select the best model out of the 4 candidates. Having decomposed every embryo into 32 trajectories as described above, we used maximum likelihood estimation to fit the parameters of every candidate model to hundreds of observations (trajectories) in the (y, G) phase space. YAP-CDX2 models were fit to $N = 9$ embryos resulting in $n = 288$ observations, and for YAP-SOX2, $N = 6$, $n = 192$. These datasets could be summarized for every stage (Fig. 3B). The dynamic nature of the data, however, allowed us to extract dependencies between the variables at adjacent stages and fit the transition matrices which fully define each model. Bayesian formalism could then be used for model selection via Bayesian Information Criterion (BIC). BIC employs the maximum likelihood approach simultaneously penalizing models for the number of parameters to prevent overfitting. Interestingly, the same network architecture (M3, Fig 3A) was selected via BIC for both TFs (Supp. Fig. S3A). This network suggests a delay of the effect of YAP localization on nuclear protein levels at the timescale of several hours (half cell cycle length, Supp. Fig. S2A) for both TFs. The selected model also includes dependence of TF level on its previous state. Thus, the winning M3 network structure allows to formulate data-driven hypotheses on the timescales of regulation of SOX2 and CDX2 by YAP.

The CPDs, or transition probability matrices, of the winning M3 network fitted to the data provide insight into the dynamics of both YAP localization and downstream TF expression. There are 1024 (y, G) trajectories that are theoretically possible, and the joint probability distribution provided by the network assigns a probability to each of them. More precisely, the probability distribution over trajectories is decomposed into the product of the initial probability distributions $p(y_0)$ and $p(G_0)$, and CPDs of the form $p(y_{i+1} \mid y_i)$ and $p(G_{i+1} \mid G_i, y_i)$ for $i = 0 \ldots 3$. Both TFs are not initially expressed, fixing their initial condition at 0, $p(G_0 = 0) = 1$. At the same time, the initial condition for y is always fixed at 1, giving $p(y_0 = 1) = 1$, and $p(y_{i+1} \mid y_i)$ describing YAP localization dynamics for later

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

stages. E.g., the transition matrix

$$p(y_1|y_0) = \begin{array}{c|cc} & y_1 & \\ y_0 & 0 & 1 \\ \hline 1 & 0.35 & 0.65 \end{array} \tag{1}$$

contains in its columns the probabilities of excluding YAP from the nucleus ($p\left(y_1 = 0 \mid y_0 = 1\right) = 0.35$) and retaining it ($p\left(y_1 = 1 \mid y_0 = 1\right) = 0.65$) at the 8- to 16-cell state division. At the 8/16 division, it is known that most cells divide asymmetrically, giving rise to one cell with nuclear and one cell with cytoplasmic YAP. The CPD in eq. (1) corresponds to 70% rate of asymmetric division which is in agreement with available data [37,38,6,39]. We obtained analogous probabilities of YAP state transitions between every two consecutive timepoints. Similarly, we fit the CPDs for every G_i node and its parents. A simple example is provided by the SOX2 transitions at the early-to-late 32-cell stage:

$$p(S_4|S_3, y_3) = \begin{array}{cc|cc} & & S_4 & \\ S_3 & y_3 & 0 & 1 \\ \hline 0 & 0 & 0.47 & 0.53 \\ 0 & 1 & 0.97 & 0.03 \\ 1 & 0 & 0.03 & 0.97 \\ 1 & 1 & 0 & 1 \end{array} \tag{2}$$

From this CPD, one can conclude that, for cells with nuclear YAP ($y_3 = 1$) and for SOX2+ cells at the 32E stage ($S_3 = 1$), the state of SOX2 propagates to the 32L stage with probability close to 1. However, for SOX2− cells with cytoplasmic YAP at the 32E stage ($S_3 = 0$, $y_3 = 0$), SOX2 gets induced by the 32L stage ($S_4 = 1$) in a non-deterministic fashion, with probability 0.53. Non-deterministic elements (transition probabilities away from 0 or 1) were also contained in CPDs at other stages for the YAP-SOX2 model as well as for some CPDs of the YAP-CDX2 model. This feature observed in both models provides another motivation for our probabilistic modeling approach.

Although some progress in probing YAP-CDX2 and YAP-SOX2 associations is possible by direct analysis of the observed trajectories, modeling can be used for inference. Bayesian modeling can be used to calculate posterior probabilities exactly or approximately, via sampling. As we illustrate below, our trained Bayesian models can be used to augment the empirical dataset by synthetic trajectories consistent with the inferred transition matrices. Such use of Bayesian models for statistical analysis of system dynamics is common in other contexts [40,41] and could be applied here to sample from the posterior distribution of any embryonic statistic.

Model-based inference

We used Bayesian networks trained on pairwise dynamic data for (y, G) to generate synthetic embryos. Every embryo in our synthetic dataset starts with 8 cells, each undergoing two rounds of division over 5 stages resulting in 32 cells at the final stage (Fig. 3C, Methods). An embryo is comprised of 8 independently sampled lineage trees, each starting with one cell sampling from the initial distribution $p(y_0, G_0) = p(y_0)p(G_0)$. The cells in each tree undergo transitions in the (y, G) phase space, sampling from the corresponding transition matrices. At divisions, expression values for daughter cells are independently sampled conditional on their mother. Every synthetic embryo could then be viewed as an ensemble of 32 trajectories

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of expression histories of cells at the 32L stage and their predecessors. These trajectories are pairwise independent for cells from different lineage trees but are not necessarily independent within one lineage. Ultimately, this approach can be applied to any dynamic Bayesian network to simulate expression dynamics of any number of variables, taking into account lineage relationships.

Since YAP is a key regulator for both CDX2 and SOX2, it is natural to ask whether particular features of YAP dynamics are a necessary or sufficient condition for induction of these TFs. Furthermore, as shown above, the expression dynamics of YAP targets in different cells can be classified by the timing of their induction. We asked whether the cells that belong to different classes could be distinguished by differences in their YAP localization time courses. If YAP were the sole activator of CDX2, it could be expected that early expression of CDX2 would be associated with more persistent levels of nuclear YAP. In contrast, one might expect that the first cells to express SOX2 should be characterized by early loss of nuclear YAP. Note that similar questions naturally arise whenever one regulator controls several targets, e.g., in the context of tissue patterning by morphogen gradients [42].

We used Bayesian inference to investigate variability in the induction timing of YAP targets, and the extent to which this variability can be explained by differences in nuclear YAP. We first generated thousands of 32L embryos describing dynamics in the (y, G) phase space, independently for $G = S$ and $G = C$. Each trajectory in an embryo belongs to one of the $G^+_{16L}, G^+_{32E}, G^+_{32L}, G^-$ classes indicating its time of induction (see **Live imaging YAP and its targets** for definition). Thus, we could estimate the expected frequency of each class for G, as well as the inter-embryonic variability of the frequencies.

Analyzing the simulated data, we found that CDX2 is most frequently induced at 16L or 32E stages (Fig. 3D), at proportions 0.64 ± 0.12 and 0.13 ± 0.8, respectively. Trained Bayesian networks were then used to predict the joint posterior distribution for y dynamics of a cell/trajectory conditional on its induction class (Fig. 3E, Methods). Interestingly, we observed little difference in the time courses of nuclear YAP in cells that induced CDX2 at the 16L and 32E stages, with both distributions mainly concentrating on cells with consistently nuclear YAP (70% and 55%, respectively). Therefore, YAP dynamics cannot be reliably used as a predictor of CDX2 induction timing (see Supp. Fig. S3C for a representative time course). In CDX2+ cells ($C_4 = 1$), 71% were predicted to have consistently nuclear YAP, with 96% of the trajectories having nuclear YAP at 16L stage (Fig. 3F), identifying nuclear YAP at this stage as a necessary condition for CDX2 expression by 32L stage. At the same time, nuclear YAP at 16L stage was predicted to result in CDX2 induction with overwhelming probability, making it a sufficient condition for CDX2 expression by 32L stage.

For SOX2, induction was predicted to be most likely at 16L or 32L stages (0.22 ± 0.10 and 0.11 ± 0.06 frequency, respectively, Fig. 3G). Conditioning on SOX2 classes of induction, we found that early loss of nuclear YAP was a necessary condition for the early onset of SOX2 expression (Fig. 3H). More precisely, all S^+_{16L} cells were predicted to have removed YAP from the nucleus at the 8/16 cell division. At the same time, most of S^+_{32L} cells were also predicted to lose YAP at one of the divisions, with probability 0.32 at 8/16 division and probability 0.59 at 16/32 cell division. Thus, loss of nuclear YAP at a division is a necessary condition for SOX2 induction. However, trajectories losing nuclear YAP at each of the divisions can fail to induce SOX2 at the next stage (Fig. 3I). Specifically, after loss of nuclear YAP at 8/16 division, cells do not induce SOX2 at 16L with probability 0.38, and after YAP loss at 16/32 division, there is no SOX2 expression with probability 0.43. In summary, loss of YAP at a division is a necessary but not sufficient condition for induction of SOX2. Most cells induce SOX2 at the second half of the subsequent cell cycle, however,

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

some cells may delay SOX2 induction or never induce it by the 32L stage (see Supp. Fig. S3D for representative examples).

Model-based data fusion

Importantly, having trained Bayesian networks for pairwise dynamics, we could establish a framework for modeling dynamics in the joint (y, C, S) phase space. Indeed, the M3 architectures of (y, C) and (y, S) models could be fused via their common Markov chain in y variable (Fig. 4A). With separate datasets for (y, C) and (y, S), we can train different parts of the network independently, accounting for missing data (Methods). The fused network was used to generate synthetic (y, C, S) lineages and simulate embryos comprised of 8 lineage trees analogously to pairwise models. Expression of every simulated lineage could be viewed as a branching trajectory in the (y, C, S) phase space, i.e., on a $\{0, 1\}^3$ cube (Supp. Fig. S4A). At every stage i, $(C_i, S_i) \in \{0, 1\}^2$ which led us to classify trajectories for every i into 4 classes (states): C^+S^- if $C_i = 1, S_i = 0$, C^-S^+ if $C_i = 0, S_i = 1$, double positive, C^+S^+, if $C_i = 1, S_i = 1$, and double negative, C^-S^-, if $C_i = 0, S_i = 0$.

We characterized the composition of a synthetic embryo at every stage i by the numbers $N_{\text{class}}(i)$ of cells in the four classes (Supp. Fig. S4B). With the number of cells fixed at every stage, the number of double negative, C^-S^-, cells can always be inferred from the rest. In other words, every simulated embryo can be described by its trajectory in the $(N_{C^+S^-}, N_{C^-S^+}, N_{C^+S^+})$ phase space. Synthesizing embryos with the fused model, we could extract the joint probability distribution of cell state proportions at every stage. In particular, we could sample from the marginal distributions at every stage (Fig. 4B). E.g., at the final stage, the fused model predicts that one can expect 19.5 ± 3.4 C^+S^- cells, 5.7 ± 2.7 C^-S^+ cells, 2.5 ± 1.6 C^-S^- cells, and 4.4 ± 2.0 C^+S^+ cells in an embryo.

Interestingly, the fused model predicted the appearance of double positive C^+S^+ cells by the 32L stage in most embryos. Furthermore, we could extract the posterior marginal distributions of y, C, and S in the C^+S^+ class (Fig. 4C). While different YAP trajectories can give rise to a C^+S^+ cell at the 32L stage, this behavior was predicted to most frequently result from loss of YAP at the 16/32 division (with $y = [1, 1, 1, 0, 0]$ with 33% probability). 60% of the C^+S^+ cells were predicted to induce CDX2 at $i = 2$, 16L stage. At the same time, we saw SOX2 induction at $i = 4$, 32L stage, with probability 55%.

To validate our findings, we imaged the joint dynamics of YAP-miRFP670, CDX2-eGFP and mScarlet-I-SOX2, along with H2B-miRFP720, in 2 embryos expressing all four reporters. Applying the same pipeline for segmentation, signal extraction and following the same discretization guidelines for each regulator as before, we identified 4 C^+S^+ cells in these embryos (see Fig. 4E for a representative example). We then compared the discretized profiles of y, C and S in these embryos (Fig. 4D) to our predictions. While the identified C^+S^+ cells indeed demonstrated variable y, C and S trajectories, more frequent trajectories for each variable coincided with the modes of the corresponding posterior marginal distributions predicted by the model in C^+S^+ cells, further validating our approach.

Discussion

The molecular components and cellular processes involved in the first cell fate bifurcation in mammalian development have been mainly elucidated based on data from fixed embryos. However, understanding the expression dynamics of this bifurcation requires access to lineages of individual blastomeres, which is impossible to obtain from fixed embryos, due to the

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

well-recognized variations in blastomere arrangements and other sources of gene expression variability at this stage. Here, we present a live imaging and cell tracking approach which reveals dynamics of nuclear YAP and its transcriptional effects across cell cycles. We also demonstrate that the time series extracted from blastomere lineages enable probabilistic modeling and model-based inferences about the dynamic process of cell allocation to the ICM and TE compartments.

Our modeling approach relies on the formalism of dynamic Bayesian networks. Bayesian networks provide a decomposition of the joint probability distribution of the nodes which can be used to calculate posterior probabilities and test specific hypotheses. Using this approach, we analyzed the necessity and sufficiency of specific features of YAP dynamics for the expression of two downstream targets, CDX2 and SOX2. Our model predicted that nuclear YAP at 16L stage is both required and sufficient for CDX2 expression, however, YAP was not a good predictor of the exact timing of expression onset. This indicates that additional modifiers are likely variable in cells that contribute to the regulation of CDX2 expression alongside YAP. A possible candidate is the Notch pathway, that has been shown to regulate *Cdx2* in parallel to Hippo signaling [43,44,45].

In the case of SOX2 we find that removal of YAP from the nucleus is required for SOX2 expression. As YAP removal mainly occurs at either the 8/16 or the 16/32 cell divisions and SOX2 expression follows with some delay, we find that SOX2 in mostly activated in two waves: at the 16L and 32L cell stages. Interestingly, however, YAP removal is not sufficient for SOX2 expression, as we find significant stochasticity in whether SOX2 is induced or not in these cells. Indeed, SOX2 was shown to be prematurely activated in only a subset of cells of 8-cell stage embryos lacking TEAD4 or YAP [19]. These observations highlight that we are lacking critical regulators of SOX2 in the embryo, either an additional repressor or an activator, whose activity is highly heterogeneous among ICM cells. The timing of ICM fate establishment has been shown to have implications for the subsequent cell fate decision, when ICM cells differentiate into epiblast and primitive endoderm cells, with early specifying ICM cells displaying a bias towards the epiblast fate [46,47,48,37,49]. SOX2, in particular, may have important roles in this process, as it is an upstream regulator of *Fgf4* expression, a key ligand produced by epiblast cells [50]. Therefore, revealing the sources of variability in the timing of ICM fate acquisition and, specifically, understanding the drivers of SOX2 induction is important for investigating the dynamics of the next cell fate decision.

Bayesian networks can account for missing data and can be used for data fusion [33]. We have demonstrated this by integrating pairwise observations into the joint description of YAP, CDX2, and SOX2. This allowed us to reveal the frequencies and origins of double positive cells in the embryo. These cells most frequently arise from early CDX2 induction, followed by loss of nuclear YAP (likely an internalization event) and subsequent induction of SOX2. This is in agreement with previous observations of CDX2 positive inner cells resulting from internalization events [24,23], which will likely completely downregulate CDX2 and assume an ICM identity. In the future, this data fusion approach can be extended to synthesize trajectories that include other important variables, such as cell shape and polarity and cell cycle progression.

Data and code availability.

Data generated and analyzed in this manuscript is available upon request. The code generating the figures is available at https://github.com/MariaAvdeeva/MouseEmbryoSimulator.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgments. We thank the members of the Developmental Dynamics Group at the Center for Computational Biology at the Flatiron Institute for helpful discussions. We thank Lucy Reading-Ikkanda (Flatiron Institute) for assistance with the graphic design of figures and Abhishek Biswas (Princeton University) for assistance with data analysis. The Flatiron Institute is a division of the Simons Foundation. Research reported in this publication was supported by the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development under award numbers R01HD110577 and R01HD107026 (to E.P.), and by the National Institute of General Medical Sciences under award number R01GM134204 (to S.Y.S.).

Author contributions. Conceptualization: M.A., S.Y.S., E.P. Methodology: B.J., M.A., M.C. Experimental data collection: M.C. Data processing and tracking: M.C. Computational modeling and analysis: M.A. Writing - original draft: M.A., S.Y.S., E.P. Writing - review & editing: M.C., M.A., S.Y.S., E.P. Supervision: S.Y.S., E.P. Project administration: M.A., E.P. Funding acquisition: E.P., S.Y.S.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods

Transgenic mouse lines. *Yap-miRFP670* (Gu 2022), *mScarlet-I-Sox2* (generated in this study), and *Cdx2-eGFP* (McDole 2011) mouse lines were used in this study. All mice were bred on a CD-1 (ICR) (Charles River) background. All animal work was carried out at Princeton University, according to the National Research Council's Guide for the Care and Use of Laboratory Animals. Animal maintenance and husbandry were carried out in accordance with the Laboratory Animal Welfare Act. All procedures were approved by Princeton University's Institutional Animal Use and Care Committee (IACUC protocol #2133). Mice were housed in a 21°C facility with a 14 hour daily light cycle and 48% average ambient humidity.

Generation and validation of mScarlet-I-Sox2 reporter mouse line. The *mScarlet-I-Sox2* targeting vector was custom synthesized (Qinglan Biotech). The targeting vector consists of a 988 bp 5' homology arm (the start codon and the genomic sequence immediately upstream of the start codon of *Sox2*), the coding sequence of mScarlet-I, a linker sequence (3 x GGGGS), and an 800 bp 3' homology arm (the genomic sequence immediately downstream of the start codon of *Sox2*) in a pMV backbone. The targeting plasmid was prepared using an endotoxin-free Maxi prep kit (Macherey-Nagel, NucleoBond Xtra Maxi EF, 740424.50). A sgRNA targeting the sequence around the start codon of *Sox2* was designed using CRISPOR (http://crispr.mit.edu/) and synthesized (Synthego). *Sox2* N-term sgRNA: CGCCCGCATGTATAACATGATGG (TGG is the PAM). For details of the targeting strategy see Supp. Fig. S1A.

Transgenic mice were generated by 2-cell cytoplasmic microinjection using the 2C-HR-CRISPR method (Gu 2018). To obtain embryos for microinjection, 5-6 week old female CD-1 (ICR) (Charles River) mice were superovulated via intraperitoneal injection with 5 IU pregnant mare serum gonadotropin (PMSG, Sigma, G4527), followed by 5 IU human chorionic gonadotropin (hCG, Sigma, 9002-61-3) 47 hours later. Superovulated females were then mated with 8-20 week old CD-1 males and embryos derived from this cross were isolated at E1.5 (2-cell) by dissecting and flushing oviducts. Microinjection of E1.5 embryos was performed in M2 media (CytoSpring, M2115) in an open glass chamber using a Leica DMi8 inverted microscope equipped with a FemtoJet (Eppendorf), micromanipulators (Leica microsystems), and a pinpoint electroporator device (micro-ePore, World Precision Instruments). Microinjection mixes were prepared in nuclease-free injection buffer (10 mM Tris-HCl, pH 7.4 and 0.25 mM EDTA) with 30 ng/μL targeting plasmid (as described above), 100 ng/μL *Cas9* mRNA and 50 ng/μL sgRNA. See below details on *Cas9* mRNA and sgRNA microinjection reagents. Microinjected embryos were immediately transferred to pseudopregnant females via surgical oviduct transfer and gestated until birth. An ear biopsy of the pups was obtained at two weeks of age, genomic DNA was isolated using and Extract-N-Amp kit (Sigma, XNAT2-1KT) and genotyping PCR was used to identify founder mice containing the *mScarlet-I-Sox2* edit. Primers used: TCCCACAACGAGGACTACAC and CTTCAGCTCCGTCTCCATCA at an annealing temperature of 60°C. Founders were then crossed to CD-1 mice to obtain the N1 generation, which were genotyped using PCR and Sanger sequenced for correct integration of the reporter construct. N1 mice identified as heterozygous were bred together, however, only heterozygous offspring were obtained, indicating that this transgenic line could only be maintained in the heterozygous state. *mScarlet-I-Sox2* heterozygous animals were fertile and did not present any obvious phenotype. For validation of mScarlet-I-SOX2 expression, embryos were treated for 48 hours,

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

from E1.5 (2-cell) through E3.5 (blastocyst) with 20uM ROCKi (Sigma, Y0503) in KSOM Embryomax (Sigma, MR-101-D) under LifeGlobal paraffin oil (CopperSurgical, LGPO-500) at 37°C, 5% O2, and 5% CO2.

Reagents for microinjection. In vitro transcription of mRNA was performed as described previously (Gu 2018). Briefly, pCS2-H2B-miRFP720 plasmid (cloned into pCS2 vector using miRFP720 sequence obtained from Addgene plasmid #136560) or pCS2-Cas9 plasmid (Addgene 122948) was linearized with NotI (New England Biolabs, R3189L) digestion and mRNA was synthesized using an mMessage mMachine SP6 intro transcription kit (Thermo Fisher Scientific, AM1340). mRNA was purified using an RNeasy Cleanup kit (Qiagen, 74104). mRNA was eluted into RNAse-free water and stored at -80°C.

Embryo isolation and microinjection for imaging. For live imaging experiments 2-cell stage (E1.5) embryos were isolated from *YAP-miRFP670*; *mScarlet-I-Sox2* females crossed with *YAP-miRFP670* males, or from *YAP-miRFP670* females crossed with *YAP-miRFP670*; *Cdx2-eGFP* males, or from *YAP-miRFP670*; *mScarlet-I-Sox2* females crossed with *YAP-miRFP670*; *Cdx2-eGFP* males after natural mating. Females were 6-15 week old and males were 8-20 week old. Oviducts were flushed with M2 media (Zenith Biotech, M2116) and embryos were washed through microdrops of M2 under LifeGlobal paraffin oil (LGPO) (CopperSurgical, LGPO-500) before transferring to pre-calibrated microdrops of KSOM EmbryoMax (Sigma, MR-101-D) under LGPO in a 37°C, 5% CO2 incubator for culture. Both cells of 2-cell stage embryos were microinjected in M2 media in an open glass chamber with 75 ng/μL H2B-miRFP720 mRNA. The same microinjection setup was used as described above for CRISPR genome editing of embryos. Injection mixes were prepared in a nuclease-free injection buffer (10 mM Tris-HCl, pH 7.4 and 0.25 mM EDTA). Following injection, embryos were cultured in KSOM under LGPO until the start of imaging.

Light sheet microscopy. Light sheet time lapse images were acquired on an inverted light sheet microscope (InVi from Luxendo/Bruker). The microscope was outfitted with an incubation chamber maintained at a constant temperature of 37°C, 5% CO2, 5% O2, and 95% relative humidity throughout the duration of the imaging. Individual embryos resided in $\sim$100μm deep wells made by gently pressing a blunt capillary tip into the base of a v-shaped chamber made of Fluorinated Ethylene Propylene (FEP) foil (model # 80-0031-02-00). Approximately six to fifteen embryos are loaded into the chamber for each experiment, each residing in its own $\sim$100μm deep well. Z-stacks for each embryo are acquired sequentially in up to four consecutive channels every 15 or 30 minutes for a total duration of approximately 30 hours. Z-stacks of YAP-miRFP670, and H2B-miRFP720 were acquired every 15 minutes, while z-stacks of mScarlet-I-SOX2 and CDX2-eGFP were acquired every 30 minutes. Images were acquired with 2.0μm z-axis resolution and 0.208μm x- and y-axis resolution. Fluorescent signal was captured using the following laser wavelengths and filters: CDX2-eGFP 488nm excitation wavelength/497-554 BP emission filter, mScarlet-I-SOX2 561nm excitation wavelength/577-612 BP emission filter; YAP-miRFP670 641nm excitation wavelength/659-690 BP emission filter; H2B-miRFP720 690nm excitation laser/700LP emission filter.

Image analysis of light sheet microscopy data. Light sheet microscopy data was analyzed following the image analysis pipeline outlined in [36]. Briefly, light sheet microscopy

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

data was first converted using lossless compression to keller-lab-block (klb) format before being segmented for nuclei using a 3D-Stardist algorithm trained on preimplantation embryos [36]. The resulting nuclear regions of interest (ROI's) were checked and manually corrected, if necessary, in AnnotatorJ. Binary nuclear masks were then generated and used to register consecutive time frames using a modified CPD-based MATLAB algorithm. Registered nuclear masks were then tracked through time in a semi-automated fashion. Average TF intensity was extracted from within the nuclear ROI's for each point in time. To correct for possible misalignment between the histone (hence nuclear masks) and TF channels, we performed rigid registration between the TF channel, aligning it to the histone channel. For that, we chose the transformation that minimizes the mean square deviation between their pixel intensities using matrix adaptation evolution strategy [51]. After alignment, TF intensity extraction was performed over each nuclear mask using the MATLAB function 'regionprops3'. For every branch, the extracted signal was smoothed over a window of 2.5h duration within every cell cycle.

Image processing for figures and videos. For visualization purposes, maximum intensity projections (MIPs) or z-slices for all figures were processed in ImageJ by first cropping the field of view to contain the embryo, then adjusting brightness contrast settings to result in optimal visual contrast. Scale bars and timestamps were added in ImageJ. Images and videos were only modified for visualization purposes; all analysis was performed on raw data.

Time warping and coarse-graining. To normalize time for each individual branch for the length of the cleavage cycles, we linearly warped the time t within every cycle between 0 and 1 ($\frac{t-\min_c t}{\max_c t-\min_c t}$ for cycle c) and rounded the times to the nearest multiple of 0.05 to form 20 timepoints per cycle. When more than one timepoint mapped to the same warped time, the average value was used. Warped times were linearly shifted according to the corresponding cleavage cycle, with [0,1], [1,2] and [2,3] corresponding to the 8, 16 and 32 cell stages respectively. To further coarse-grain time, we also used warped time to define early (E) and late (L) cycle stages by applying $x < 0.5$ and $x >= 0.5$ cutoffs on the warped times, respectively. This was applied to 16- and 32-cell stages only. As a result 5 stages were defined: 8, 16E, 16L, 32E, 32L.

Defining y, S and C via discretization. We coarse-grained YAP, SOX2 and CDX2 dynamics arriving at binary sequences y, S and C indexed by stages, $i = 0 \dots 4$. First 2 hours post division were excluded from every branch, to normalize for small perturbations introduced by cell division. For YAP, we summarized dynamics for every branch using its average (after z-scoring) over each stage, and for CDX2 and SOX2, we used their final values (after normalization) for every stage for every branch. For every stage, we then classified the branches into low (0) and high (1) expression based on these summaries. We restricted classification to 16E, 16L, 32E and 32L stages for YAP, and 16L, 32E and 32L stages for the downstream TFs. In particular, at 8-cell stage, i.e., for $i = 0$, YAP is nuclear in all cells, and we assigned $y_0 = 1$. At the same time, both TFs are not initially expressed, and we put $S_0 = S_1 = 0$ and $C_0 = C_1 = 0$. For every variable v, we classified branches for every stage i using thresholding on its stage-dependent summarized levels $\widetilde{v_{b,i}}$, with b indexing the branches. The thresholds were chosen in a data-driven manner, universally for all embryos. In particular, for YAP the thresholds were chosen to be stage-dependent; for both TFs, the thresholds were chosen to be constant over stages. See below for details of discretization

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

procedures and thresholds for each individual variable. Ultimately, for stage i and variable v, a branch b with expression above the corresponding threshold θ_i, i.e., satisfying $\tilde{v}_{b,i} > \theta_i$, was assigned $v_{b,i} = 1$. Otherwise we put $v_{b,i} = 0$.

Discretization for nuclear YAP. In every embryo, we first z-scored YAP nuclear concentration at every timepoint, to normalize for the downward trends in its average nuclear concentration. Distributions of the summary variable $\widetilde{y_{b,i}}$, i.e., average z-scored concentration at stage i where b indexes the branches, were analyzed separately for every i. We found that, at stages 16E and 16L, average YAP values exhibited bimodal distribution (Supp. Fig. S2B). To separate the modes into low and high, we extracted a kernel density estimate for each distribution using *distplot* method from the *seaborn* Python package, with default parameters, and identified its local minimum (via *scipy.signal.argrelmin*). For stages 32E and 32L, with no obvious bimodal structure present, we applied the same thresholding procedure as for CDX2 (see below). Thus, individual thresholds were applied to each half stage; $\theta_1 = -0.59$ for 16E, $\theta_2 = -0.82$ for 16L, $\theta_3 = -0.61$ for 32E, and $\theta_4 = -0.58$ for 32L.

Discretization for CDX2. For the analysis on CDX2 expression, total CDX2 signal in each nucleus was used. To classify individual branches using their CDX2 expression profiles (Supp. Fig. S2C), we first applied additional small background correction by subtracting the minimal expression value from all the data at every timepoint. We also minmax normalized total CDX2 signal in each embryo using its minimal and maximal expression at the 16- and 32-cell stages. For CDX2, we used clustering on summary data from all relevant stages (16L, 32E, 32L) to choose a universal thresholding parameter. In particular, we applied *GaussianMixture* method from the *sklearn.mixture* Python package to cluster the data into three clusters. Ordering the clusters by their mean expression, we annotated the bottom cluster as 0 (no CDX2 expression) and top two clusters (intermediate and high CDX2 expression) were annotated as 1 which is equivalent to using $\theta_2 = \theta_3 = \theta_4 = 0.088$.

Discretization for SOX2. Total SOX2 signal in each nucleus was minmax normalized and background corrected analogously to the analysis on CDX2. To arrive at the summary variables $\widetilde{S_{b,i}}$ for every branch b, normalizing for slight branch-dependent technical differences, we take the final total SOX2 level for each stage i, and subtract the total SOX2 intensity for this branch at the beginning of the 16-cell stage. For SOX2, we chose a universal thresholding parameter $\theta_2 = \theta_3 = \theta_4 = 0.08$ (Supp. Fig. S2D).

Bayesian modeling. We modeled the discretized data via Bayesian networks. To this end, we applied methods from the *BayesianNetwork* class in the *pgmpy* Python package. Nodes and edges were specified as shown in Fig. 3A for the datasets with pairwise observations and Fig. 4A for the datasets with triple (y, S, C) observations. Let us denote the set of all variables/nodes of the network as V; for a node $v \in V$ we denote the set of its parent nodes as $\mathcal{P}(v)$. Bayesian network provides a decomposition of the joint probability distribution of V using the *conditional probability distributions* (CPDs): $p(V) = \prod_v p\left(v|\mathcal{P}(v)\right)$. Here $p\left(v|\mathcal{P}(v)\right) = p(v)$ if a node has no parents. The nodes of the dynamic Bayesian networks are indexed by time i; let us denote the set of variables corresponding to the timepoint i (time slice) by V_i. For the winning model M3 that we considered in this paper (Fig. 3A), for a variable $v \in V_i$, $\mathcal{P}(v) = V_{i-1}$ and the CPD corresponding to v can be written in the

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

form $p\left(v|\mathcal{P}(v)\right) = p(v|V_{i-1})$. We will make use of this structure of the decomposition when generating synthetic embryos.

The candidate networks were fit to data using the maximum likelihood estimator, i.e., for a model M and observed dataset X, $\hat{\theta} = \text{argmax}_\theta\, p(X|\theta, M)$ was used as the optimal parameter after training. To compare Bayesian models, we computed Bayesian Information Criterion (BIC) for each model via the *structure_score* method from the *pgmpy* Python package. If n denotes the number of observations in the dataset X, and M has k parameters, BIC is computed via the following formula: $\text{BIC} = k \ln n - 2 \ln p(X|\hat{\theta}, M)$. Here $\ln p(X|\hat{\theta}, M)$ is the loglikelihood of the model after training.

Generating a synthetic embryo. We generated synthetic embryos incorporating the lineage structure; in this paper, $N = 3000$ embryos were used for every experiment. For every embryo, 8 lineages were sampled independently. To simulate a lineage, we made use of the CPDs of the trained model. Recall that the set of variables corresponding to the timepoint i was denoted by V_i. For $i \geq 1$ and $v \in V_i$, the CPDs of the M3 model have the form of the transition matrices $p(v|V_{i-1})$; for $i = 0$, the nodes have no parents and the CPDs correspond to the initial probability distributions $p(v)$, $v \in V_0$.

A simulated lineage is a collection of expression profiles of 13 synthetic cells, x_0^a, $x_1^{\{aa,ab\}}$, $x_2^{\{aaa,aba\}}$, $x_3^{\{aaaa,aaab,abaa,abab\}}$, $x_4^{\{aaaaa,aaaba,abaaa,ababa\}}$. Here the (redundant) subscript corresponds to the timepoint, and the superscript encodes the lineage structure of the tree (Supp. Fig. S3B), with the cell x_i^s encoded by the binary sequence s at timepoint i giving rise to the cells x_{i+1}^{sa} and x_{i+1}^{sb} in case of a division and the cell x_{i+1}^{sa} otherwise. sa and sb denote concatenation of the sequence s with a and b, respectively; let us denote the truncated sequence s, with the last element removed, by s'. Two divisions are happening, one between timepoints $i = 0$ and $i = 1$ and the other between $i = 2$ and $i = 3$. The expression for the cells was independently sampled from the corresponding CPDs. More precisely, we sampled x_0^a from $p(V_0)$, and sequentially sampled expression for the cells in all the subsequent timepoints, with expression for the cell x_i^s sampled from $p\left(V_i|x_{i-1}^{s'}\right)$ for $i \geq 1$. Sampling from the CPDs was realized using the *simulate* method available for Bayesian networks in *pgmpy*. Note that this strategy is appropriate for any number of variables observed over time, and only takes advantage of the fact that the parents of the nodes in our dynamic Bayesian networks belong to the previous time slice. In particular, the same simulation strategy was applied for the pairwise and fused models.

Data fusion via Bayesian modeling. To fuse the pairwise datasets corresponding to (y, S) and (y, C), we trained a dynamic Bayesian network on the concatenated data (Fig. 4A). Training in *pgmpy* allows for missing data. Due to the structure of the fused network, training it on both datasets results in the update of the CPDs for y, i.e., $p(y_0)$, $p(y_i|y_{i-1})$, $i \geq 1$, with the CPDs for TFs, i.e., $p(G_0)$, $p(G_i|y_{i-1}, G_{i-1})$, $i \geq 1$, for $G = S, C$ coinciding with the CPDs for the models only trained on paired observations.

Exact inference. Bayesian networks can be used for exact inference to compute posterior distributions of the form $p(Q|E = e)$ for a set of query variables Q given some observed evidence $E = e$. This task can be approached by algorithms such as variable elimination in which all the non-query and non-evidence variables (i.e., $V \setminus \{Q \cup E\}$ where V is the set of all variables in the model) are efficiently marginalized out in a pre-specified order. For exact

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

inference, we applied the *VariableElimination* method from *pgmpy* to corresponding trained models. In particular, for variable elimination conditional on induction classes in Figs. 3E,H we used $\{G_2 = 1\}$, $\{G_2 = 0, G_3 = 1\}$, $\{G_2 = 0, G_3 = 0, G_4 = 1\}$ as evidence for classes G^+_{16L}, G^+_{32E}, and G^+_{32L}, respectively. Analogously, for Fig. 3F, $\{C_4 = 1\}$ and $\{Y_2 = 1\}$ were used as evidence; for Fig. 3I, $\{Y_2 = 0\}$ and $\{Y_3 = 1, Y_4 = 0\}$ were used.

References

1. E. H. Davidson, "How embryos work: a comparative view of diverse modes of cell fate specification," *Development*, vol. 108, no. 3, pp. 365–389, 1990.
2. M. Zhu and M. Zernicka-Goetz, "Principles of Self-Organization of the Mammalian Embryo," *Cell*, vol. 183, no. 6, pp. 1467–1478, 2020.
3. H. T. Zhang and T. Hiiragi, "Symmetry Breaking in the Mammalian Embryo.," *Annual review of cell and developmental biology*, vol. 34, pp. 405–426, oct 2018.
4. K. Cockburn and J. Rossant, "Making the blastocyst: lessons from the mouse.," *The Journal of clinical investigation*, vol. 120, pp. 995–1003, apr 2010.
5. N. Dard, S. Louvet-Vallée, and B. Maro, "Orientation of mitotic spindles during the 8- to 16-cell stage transition in mouse embryos.," *PloS one*, vol. 4, p. e8171, dec 2009.
6. T. Watanabe, J. S. Biggins, N. B. Tannan, and S. Srinivas, "Limited predictive value of blastomere angle of division in trophectoderm and inner cell mass specification," *Development*, vol. 141, no. 11, pp. 2279–2288, 2014.
7. E. Korotkevich, R. Niwayama, A. Courtois, S. Friese, N. Berger, F. Buchholz, and T. Hiiragi, "The Apical Domain Is Required and Sufficient for the First Lineage Segregation in the Mouse Embryo.," *Developmental cell*, vol. 40, pp. 235–247.e7, feb 2017.
8. R. Niwayama, P. Moghe, Y.-J. Liu, D. Fabrèges, F. Buchholz, M. Piel, and T. Hiiragi, "A Tug-of-War between Cell Shape and Polarity Controls Division Orientation to Ensure Robust Patterning in the Mouse Blastocyst," *Developmental Cell*, vol. 51, no. 5, pp. 564–574.e6, 2019.
9. D. Fabrèges, B. Corominas-Murtra, P. Moghe, A. Kickuth, T. Ichikawa, C. Iwatani, T. Tsukiyama, N. Daniel, J. Gering, A. Stokkermans, A. Wolny, A. Kreshuk, V. Duranthon, V. Uhlmann, E. Hannezo, and T. Hiiragi, "Temporal variability and cell mechanics control robustness in mammalian embryogenesis," *Science*, vol. 386, no. 6718, p. eadh1145, 2024.
10. C. Royer, K. Leonavicius, A. Kip, D. Fortin, K. Nandi, A. Vincent, C. Jones, T. Child, K. Coward, C. Graham, and S. Srinivas, "Establishment of a relationship between blastomere geometry and YAP localisation during compaction," *Development*, vol. 147, p. dev189449, oct 2020.
11. M. H. Johnson and C. A. Ziomek, "The foundation of two distinct cell lineages within the mouse morula," *Cell*, vol. 24, no. 1, pp. 71–80, 1981.
12. A. K. Tarkowski and J. Wróblewska, "Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage.," *Journal of embryology and experimental morphology*, vol. 18, pp. 155–180, aug 1967.
13. W. R. Holmes, N. S. Reyes de Mochel, Q. Wang, H. Du, T. Peng, M. Chiang, O. Cinquin, K. Cho, and Q. Nie, "Gene Expression Noise Enhances Robust Organization of the Early Mammalian Blastocyst.," *PLoS computational biology*, vol. 13, p. e1005320, jan 2017.
14. N. Saiz, L. Mora-Bitria, S. Rahman, H. George, J. P. Herder, J. Garcia-Ojalvo, and A.-K. Hadjantonakis, "Growth-factor-mediated coupling between lineage size and cell fate choice underlies robustness of mammalian development.," *eLife*, vol. 9, jul 2020.
15. E. Wicklow, S. Blij, T. Frum, Y. Hirate, R. A. Lang, H. Sasaki, and A. Ralston, "HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst.," *PLoS genetics*, vol. 10, p. e1004618, oct 2014.
16. D. Strumpf, C.-A. Mao, Y. Yamanaka, A. Ralston, K. Chawengsaksophak, F. Beck, and J. Rossant, "Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst," *Development*, vol. 132, no. 9, pp. 2093–2102, 2005.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

17. N. Nishioka, K.-i. Inoue, K. Adachi, H. Kiyonari, M. Ota, A. Ralston, N. Yabuta, S. Hirahara, R. O. Stephenson, N. Ogonuki, R. Makita, H. Kurihara, E. M. Morin-Kensicki, H. Nojima, J. Rossant, K. Nakao, H. Niwa, and H. Sasaki, "The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass," *Developmental Cell*, vol. 16, no. 3, pp. 398–410, 2009.
18. T. Frum, T. M. Murphy, and A. Ralston, "HIPPO signaling resolves embryonic cell fate conflicts during establishment of pluripotency in vivo," *eLife*, vol. 7, p. e42298, dec 2018.
19. T. Frum, J. L. Watts, and A. Ralston, "TEAD4, YAP1 and WWTR1 prevent the premature onset of pluripotency prior to the 16-cell stage," *Development*, vol. 146, no. 17, p. dev179861, 2019.
20. Y. Hirate, S. Hirahara, K.-i. Inoue, A. Suzuki, V. Alarcon, K. Akimoto, T. Hirai, T. Hara, M. Adachi, K. Chida, S. Ohno, Y. Marikawa, K. Nakao, A. Shimono, and H. Sasaki, "Polarity-Dependent Distribution of Angiomotin Localizes Hippo Signaling in Preimplantation Embryos," *Current Biology*, vol. 23, pp. 1181–1194, jul 2013.
21. C. Y. Leung and M. Zernicka-Goetz, "Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms," *Nature Communications*, vol. 4, no. 1, p. 2251, 2013.
22. E. Posfai, S. Petropoulos, F. R. O. de Barros, J. P. Schell, I. Jurisica, R. Sandberg, F. Lanner, and J. Rossant, "Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo," *eLife*, vol. 6, p. e22906, feb 2017.
23. K. McDole and Y. Zheng, "Generation and live imaging of an endogenous Cdx2 reporter mouse line," *genesis*, vol. 50, no. 10, pp. 775–782, 2012.
24. Y. Toyooka, S. Oka, and T. Fujimori, "Early preimplantation cells expressing Cdx2 exhibit plasticity of specification to TE and ICM lineages through positional changes.," *Developmental biology*, vol. 411, pp. 50–60, mar 2016.
25. B. Gu, B. Bradshaw, M. Zhu, Y. Sun, S. Hopyan, and J. Rossant, "Live imaging YAP signalling in mouse embryo development.," *Open biology*, vol. 12, p. 210335, jan 2022.
26. P. Krupinski, V. Chickarmane, and C. Peterson, "Simulating the mammalian blastocyst–molecular and mechanical interactions pattern the embryo.," *PLoS computational biology*, vol. 7, p. e1001128, may 2011.
27. J. De Caluwé, A. Tosenberger, D. Gonze, and G. Dupont, "Signalling-modulated gene regulatory networks in early mammalian development," *Journal of Theoretical Biology*, vol. 463, pp. 56–66, 2019.
28. S. B. Nissen, M. Perera, J. M. Gonzalez, S. M. Morgani, M. H. Jensen, K. Sneppen, J. M. Brickman, and A. Trusina, "Four simple rules that are sufficient to generate the mammalian blastocyst.," *PLoS biology*, vol. 15, p. e2000737, jul 2017.
29. Z. Cang, Y. Wang, Q. Wang, K. W. Y. Cho, W. Holmes, and Q. Nie, "A multiscale model via single-cell transcriptomics reveals robust patterning mechanisms during early mammalian embryo development.," *PLoS computational biology*, vol. 17, p. e1008571, mar 2021.
30. M. A. Ramirez Sierra and T. R. Sokolowski, "AI-powered simulation-based inference of a genuinely spatial-stochastic gene regulation model of early mouse embryogenesis," *PLOS Computational Biology*, vol. 20, no. 11, pp. 1–60, 2024.
31. S. Y. Kim, S. Imoto, and S. Miyano, "Inferring gene networks from time series microarray data using dynamic Bayesian networks.," *Briefings in bioinformatics*, vol. 4, pp. 228–235, sep 2003.
32. J. Robinson and A. Hartemink, "Non-stationary dynamic Bayesian networks," in *Advances in Neural Information Processing Systems* (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), vol. 21, Curran Associates, Inc., 2008.
33. D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.
34. D. Ruiz-Perez, J. Lugo-Martinez, N. Bourguignon, K. Mathee, B. Lerner, Z. Bar-Joseph, and G. Narasimhan, "Dynamic Bayesian Networks for Integrating Multi-omics Time Series Microbiome Data," *mSystems*, vol. 6, no. 2, pp. 10.1128/msystems.01105–20, 2021.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

35. P. Suter, J. Kuipers, and N. Beerenwinkel, "Discovering gene regulatory networks of multiple phenotypic groups using dynamic Bayesian networks," *Briefings in Bioinformatics*, vol. 23, no. 4, p. bbac219, 2022.
36. H. Nunley, B. Shao, D. Denberg, P. Grover, J. Singh, M. Avdeeva, B. Joyce, R. Kim-Yip, A. Kohrman, A. Biswas, A. Watters, Z. Gal, A. Kickuth, M. Chalifoux, S. Y. Shvartsman, L. M. Brown, and E. Posfai, "Nuclear instance segmentation and tracking for preimplantation mouse embryos," *Development*, vol. 151, p. dev202817, Nov. 2024.
37. Y. Yamanaka, F. Lanner, and J. Rossant, "FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst," *Development*, vol. 137, pp. 715–724, mar 2010.
38. S. Anani, S. Bhat, N. Honma-Yamanaka, D. Krawchuk, and Y. Yamanaka, "Initiation of Hippo signaling is linked to polarity rather than to cell position in the pre-implantation mouse embryo.," *Development (Cambridge, England)*, vol. 141, pp. 2813–2824, jul 2014.
39. K. McDole, Y. Xiong, P. A. Iglesias, and Y. Zheng, "Lineage mapping the pre-implantation mouse embryo by two-photon microscopy, new insights into the segregation of cell fates," *Developmental Biology*, vol. 355, no. 2, pp. 239–249, 2011.
40. R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau, "Bayesian statistics and modelling," *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 1, 2021.
41. T. L. Griffiths, N. Chater, and J. B. Tenenbaum, *Bayesian Models of Cognition: Reverse Engineering the Mind*. Cambridge, MA: MIT Press, 2024.
42. H. L. Ashe and J. Briscoe, "The interpretation of morphogen gradients," *Development*, vol. 133, no. 3, pp. 385–394, 2006.
43. Y. Watanabe, K. Y. Miyasaka, A. Kubo, Y. S. Kida, O. Nakagawa, Y. Hirate, H. Sasaki, and T. Ogura, "Notch and Hippo signaling converge on Strawberry Notch 1 (Sbno1) to synergistically activate Cdx2 during specification of the trophectoderm," *Scientific Reports*, vol. 7, no. 1, p. 46135, 2017.
44. T. Rayon, S. Menchero, A. Nieto, P. Xenopoulos, M. Crespo, K. Cockburn, S. Cañon, H. Sasaki, A.-K. Hadjantonakis, J. L. de la Pompa, J. Rossant, and M. Manzanares, "Notch and hippo converge on Cdx2 to specify the trophectoderm lineage in the mouse blastocyst.," *Developmental cell*, vol. 30, pp. 410–422, aug 2014.
45. S. Menchero, I. Rollan, A. Lopez-Izquierdo, M. J. Andreu, J. de Aja, M. Kang, J. Adan, R. Benedito, T. Rayon, A.-K. Hadjantonakis, and M. Manzanares, "Transitions in cell potency during early mouse development are driven by Notch," *eLife*, vol. 8, p. e42930, apr 2019.
46. A. I. Mihajlović, V. Thamodaran, and A. W. Bruce, "The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent," *Scientific Reports*, vol. 5, no. 1, p. 15034, 2015.
47. S. A. Morris, S. J. L. Graham, A. Jedrusik, and M. Zernicka-Goetz, "The differential response to Fgf signalling in cells internalized at different times influences lineage segregation in preimplantation mouse embryos.," *Open biology*, vol. 3, p. 130104, nov 2013.
48. M. Krupa, E. Mazur, K. Szczepańska, K. Filimonow, M. Maleszewski, and A. Suwińska, "Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8→16- and 16→32-cells).," *Developmental biology*, vol. 385, pp. 136–148, jan 2014.
49. S. A. Morris, R. T. Y. Teo, H. Li, P. Robson, D. M. Glover, and M. Zernicka-Goetz, "Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 6364–6369, apr 2010.
50. T. K. Mistri, W. Arindrarto, W. P. Ng, C. Wang, L. H. Lim, L. Sun, I. Chambers, T. Wohland, and P. Robson, "Dynamic changes in Sox2 spatio-temporal expression promote the second cell fate decision through Fgf4/Fgfr2 signaling in preimplantation mouse embryos.," *The Biochemical journal*, vol. 475, pp. 1075–1089, mar 2018.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

51. H.-G. Beyer and B. Sendhoff, "Simplify your covariance matrix adaptation evolution strategy," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 5, pp. 746–759, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

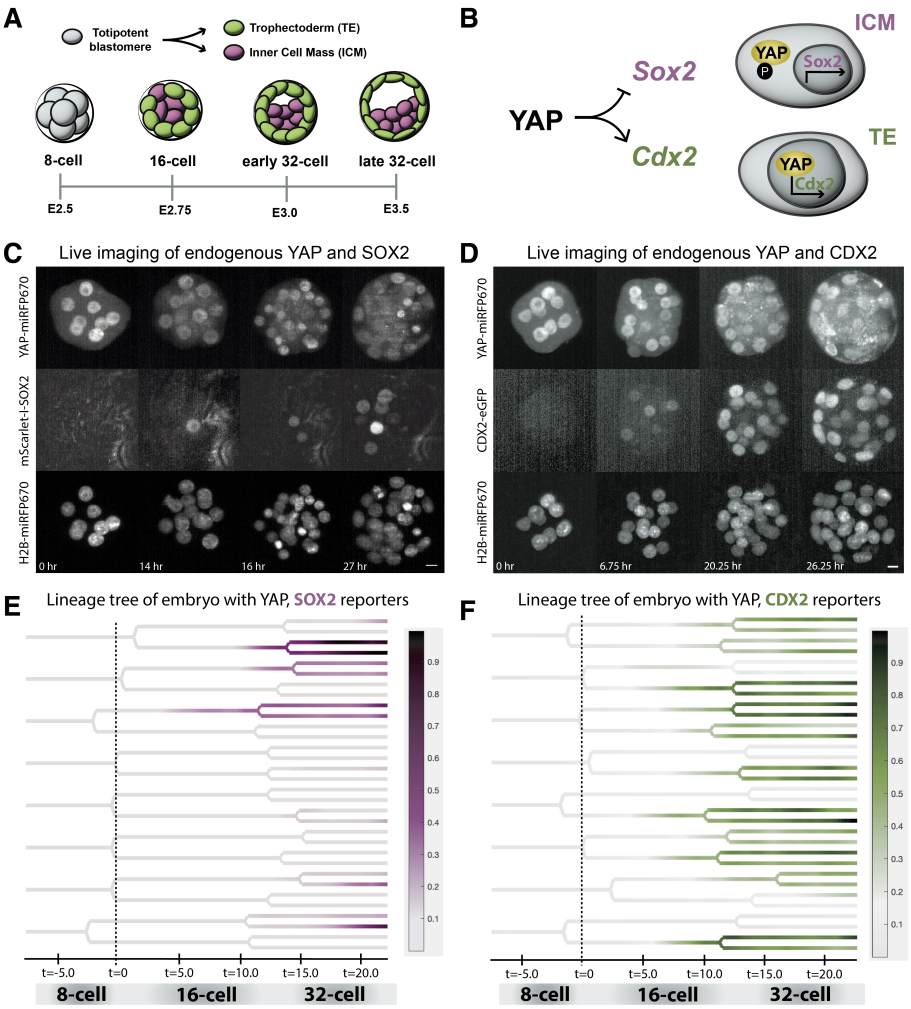


Fig. 1: Live imaging of endogenously tagged YAP-miRFP670, mScarlet-I-SOX2, and CDX2-eGFP during the first fate decision in preimplantation embryos. (A) A developmental timeline of TE and ICM specification. The first cell differentiation event, between TE (green) and ICM (purple), initiates at the 16-cell stage, establishing distinct inner, ICM, and outer, TE, cell populations in the 32-cell stage embryo. Embryos in this study are imaged through the 32-cell stage. Days post-fertilization are denoted by embryonic day (E). **(B)** *Sox2* and *Cdx2* expression is regulated by YAP. In ICM cells, YAP is phosphorylated and retained in the cytoplasm, allowing for *Sox2* expression. In TE cells, YAP is predominantly localized to the nucleus where it activates expression of *Cdx2*. **(C)** Time lapse images of a representative embryo, expressing YAP-miRFP670, mScarlet-I-SOX2, and H2B-miRFP670, from the 8-cell through the late 32-cell stage. A z-stack was acquired every 15 minutes for YAP and H2B channels and every 30 minutes for the SOX2 channel. Maximum intensity projections (MIPs) are shown. Unit of time is hours. Scale bar: 10μm. **(D)** Same as **(C)** for a representative embryo, expressing YAP-miRFP670, CDX2-eGFP, and H2B-miRFP670, from the 8-cell through the late 32-cell stage. **(E)** Lineage trees from an embryo expressing YAP-miRFP670 and mScarlet-I-SOX2, tracked until the first cell division into the 33-cell stage. Expression of mScarlet-I-SOX2 intensity (minmax normalized from 0 to 1) is colormapped in purple onto the branches of the lineage tree. t=0 is defined as the average time of division from the 8-16 cell stage. Unit of time is hours. **(F)** Same as **(E)** for an embryo expressing YAP-miRFP670 and CDX2-eGFP, tracked until the first cell division into the 33-cell stage. Minmax normalized CDX2-eGFP intensity is colormapped in green.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

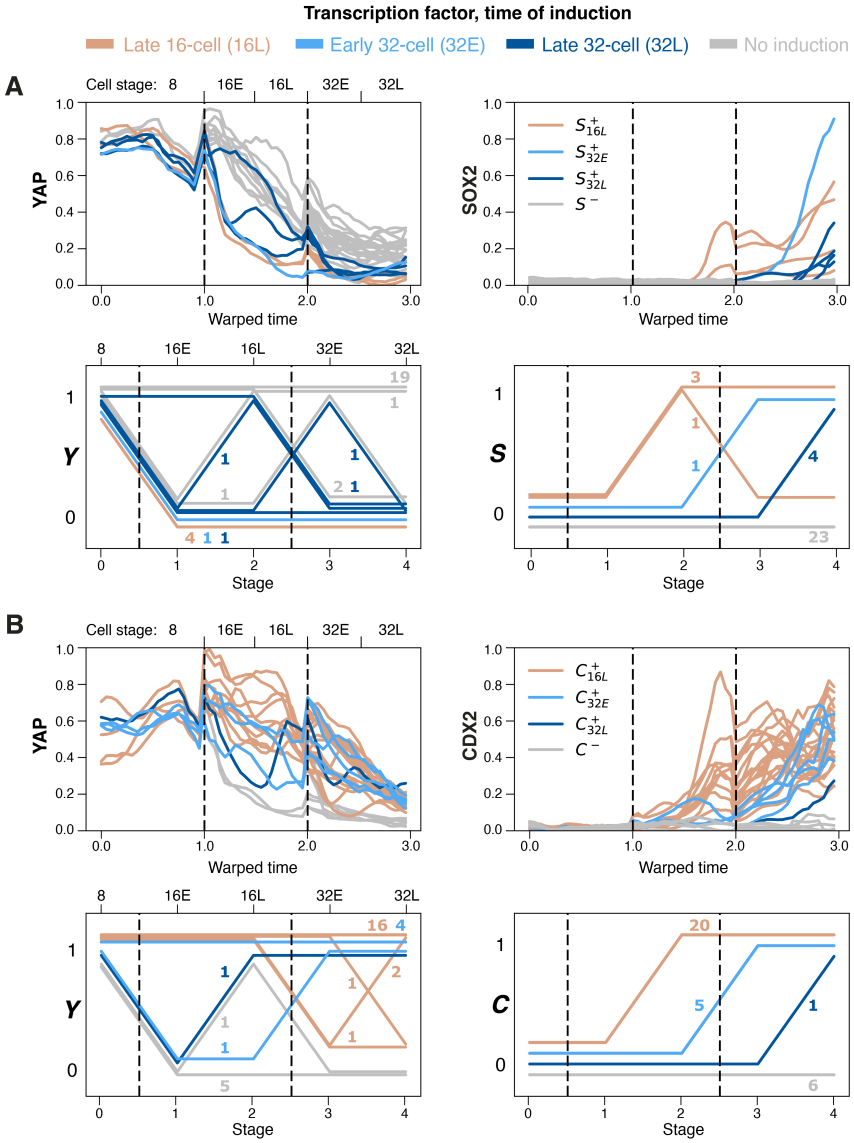


Fig. 2: Dynamics of nuclear YAP concentration and expression of SOX2 and CDX2. **(A)** Extracted and discretized dynamics of YAP localization and SOX2 expression in the cell lineages of one representative embryo. Imaging starts at the 8-cell stage and all cells undergo 2 division rounds. Color: each branch is colored according to its induction class (S^+_{16L}, S^+_{32E}, S^+_{32L}, S^-), i.e., stage of SOX2 induction. See below for the description of stages. Top left and right, x-axis: linearly warped time, with division times aligned for all branches. Top left, y-axis: average YAP reporter intensity in segmented nuclei, smoothed and minmax normalized between 0 and 1. Top right, y-axis: total SOX2 reporter intensity in each segmented nucleus, smoothed and minmax normalized between 0 and 1. Bottom left and right, x-axis: 5 stages, 0(8): 8-cell stage, 1(16E): first half of 16-cell stage, 2(16L): second half of 16-cell stage, 3(32E): first half of 32-cell stage, 4(32L): second half of 32-cell stage. Black dashed vertical lines indicate 8/16 cell and 16/32 cell division times. Bottom left, y-axis: same as top left, where YAP is averaged over every stage and discretized (Methods). Bottom right, y-axis: same as bottom left but for discretized total SOX2 reporter intensity. 32 branches of YAP localization histories are shown, grouped by their YAP behavior and induction class. Number of branches in each group is enumerated next to the discretized trajectory. **(B)** Same as **(A)** for YAP and CDX2 in a different representative embryo. Each branch is colored according to its stage of CDX2 induction.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

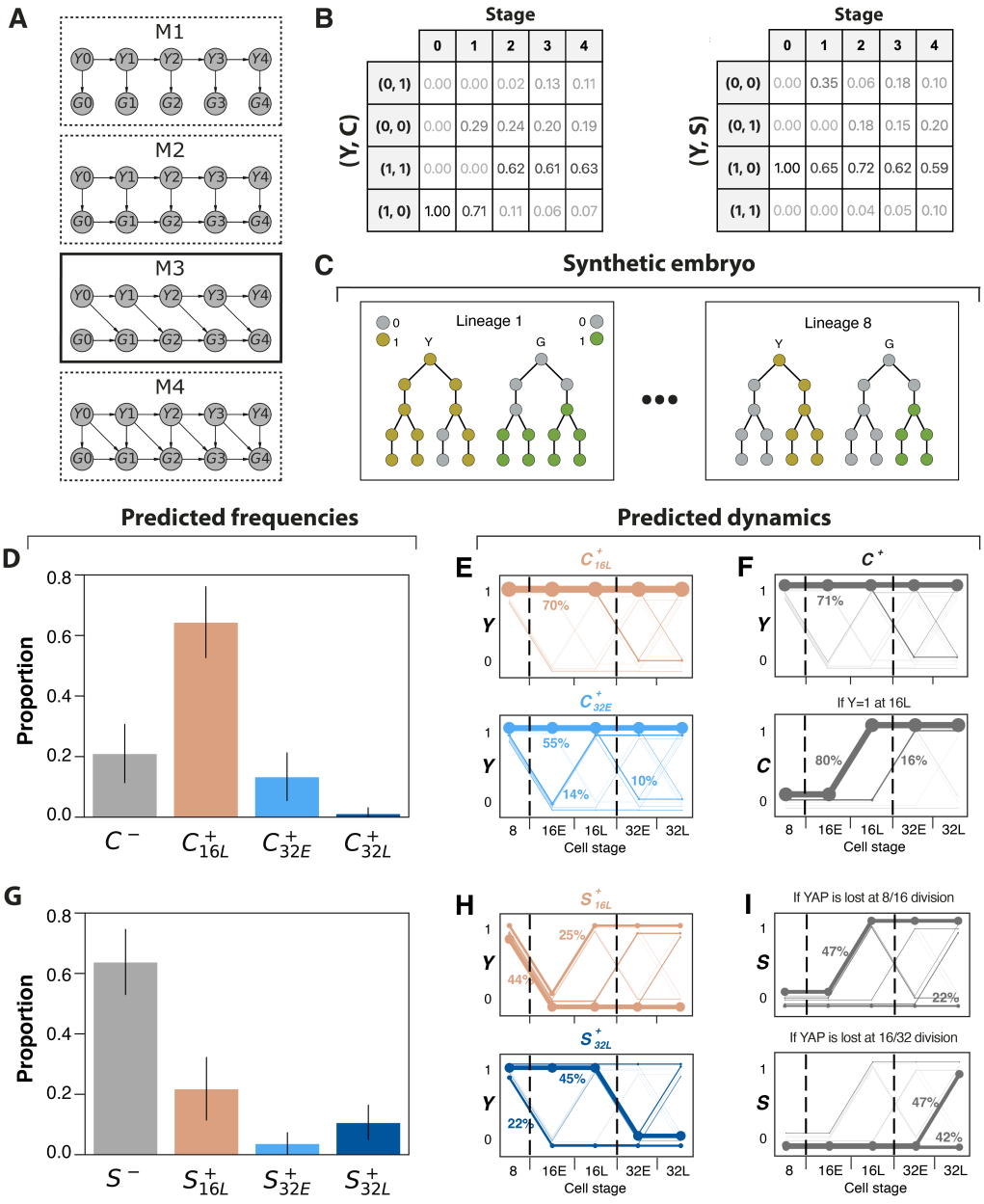


Fig. 3: Caption on the next page.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 3: Modeling gradual induction of CDX2 and SOX2 with dynamic Bayesian networks. (A) Bayesian network architectures considered. Y_i: binarized nuclear YAP concentration at stage $i = 0 \dots 4$. G_i: total nuclear level of a TF at stage i. $G = C$ for CDX2 and $G = S$ for SOX2. The winning M3 model is shown with the solid outline. **(B)** Table summarizing frequencies of binarized (y, G) observations at every stage. Left: CDX2, $N = 9$ embryos. Right: SOX2, $N = 6$ embryos. 8 observations (cells) per embryo for $i = 0$; 16 cells per embryo for $i = 1, 2$; 32 cells per embryo for $i = 3, 4$. **(C)** A schematic of an embryo simulated from a Bayesian network (Methods). A synthetic embryo is a collection of 8 independent lineage trees (sample trees for lineages 1 and 8 are shown). Every lineage is comprised of 13 synthetic cells: one cell undergoing 2 rounds of division over 5 stages. For a lineage, left: nuclear YAP ($y = 1$) is shown in olive, cytoplasmic YAP ($y = 0$) in grey. Right: gene expressed ($G = 1$) is shown in green, not expressed ($G = 0$) in grey. **(D)** Frequencies of CDX2 induction classes over 3000 simulated embryos. Bar: mean frequency, error bar: one standard deviation around the mean. **(E)** Top: posterior distribution of y for cells inducing CDX2 at 16L stage (i.e., conditioned on $C_2 = 1$). Linewidth of a trajectory is proportional to its posterior probability. Posterior probabilities are marked next to the corresponding, most prominent branches. Bottom: same for induction at 32E stage. **(F)** Top: same as **(E)** for all CDX2+ cells ($C_4 = 1$). Bottom: posterior distribution of C trajectory for cells with nuclear YAP at 16L stage (i.e., conditioned on $Y_2 = 1$). **(G)** Same as **(D)** for SOX2. **(H)** Same as **(E)** for cells inducing SOX2 at 16L stage (top) or inducing SOX2 at 32L stage (bottom). **(I)** Top: posterior distribution of S for cells losing nuclear YAP at 8/16 cell division (i.e., conditioned on $Y_1 = 0$). Linewidth of a trajectory is proportional to its posterior probability. Posterior probabilities are marked next to the corresponding, most prominent branches. Bottom: same for cells losing nuclear YAP at 16/32 cell division ($Y_2 = 1, Y_3 = 0$).

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

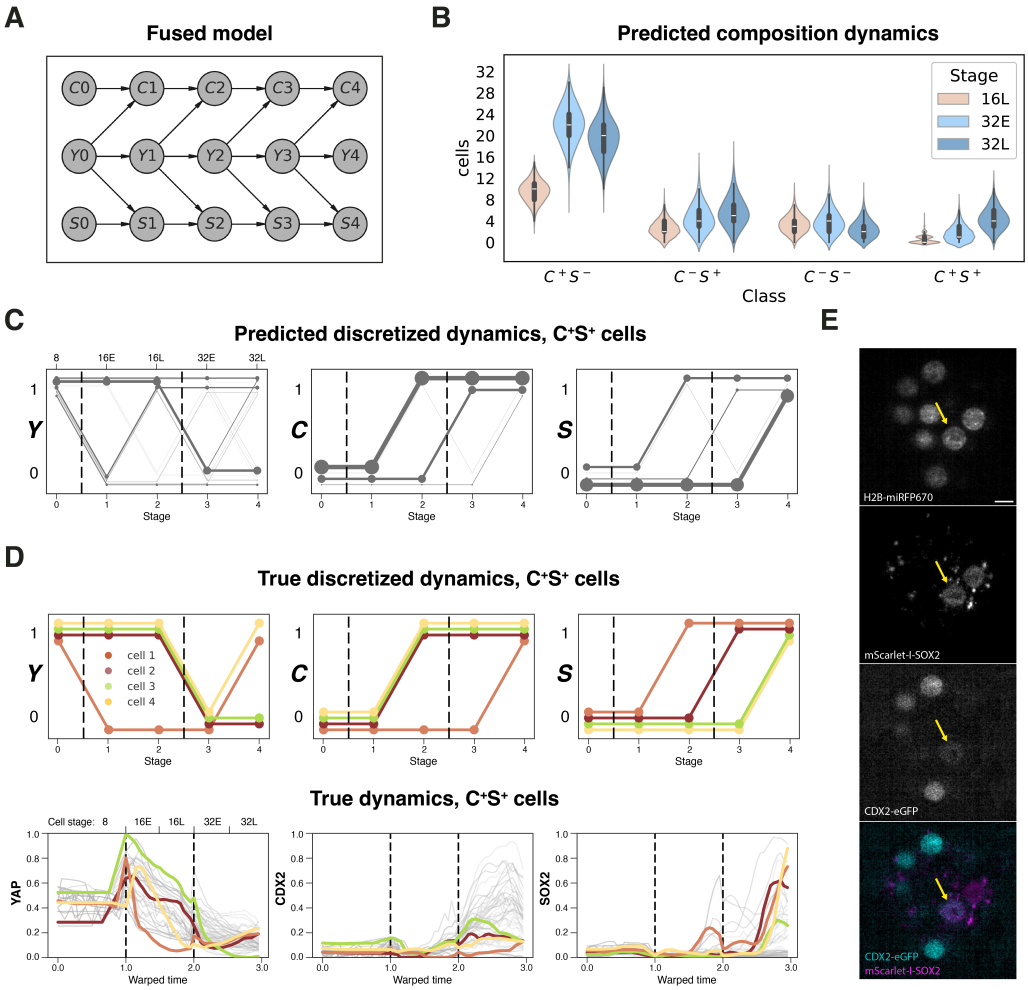


Fig. 4: Modeling joint dynamics of YAP, CDX2 and SOX2 with a fused Bayesian network. (A) The structure of the network that was used for data fusion. C_i: (binarized) nuclear CDX2 level at stage i, S_i: (binarized) nuclear SOX2 level at stage i, Y_i: (binarized) nuclear YAP concentration at stage i. **(B)** Violin plot showing posterior marginal distributions for the numbers of cells in four possible classes at 16L, 32E and 32L stages. Distributions are smoothed with kernel density estimation. 3000 embryos were simulated. **(C)** Left: posterior distribution of y for C^+S^+ cells (i.e., conditioned on $C_4 = 1, S_4 = 1$). Linewidth of a trajectory is proportional to its posterior probability. Middle: same for C. Right: same for S. **(D)** Top: Discretized trajectories of Y, S and C for 4 C^+S^+ cells identified in 2 embryos simultaneously expressing YAP-miRFP670, CDX2-eGFP and mScarlet-I-SOX2. Bottom: nuclear YAP concentration (left), total nuclear levels of CDX2 (middle) and total nuclear levels of SOX2 (right) for the same 4 cells as in top. Data from all other cells in the 2 embryos are shown in gray. Values on y-axis were smoothed and minmax normalized between 0 and 1 within every embryo. x-axis: linearly warped time, with division times aligned. Black dashed vertical lines: 8/16 cell and 16/32 cell division times. **(E)** Light sheet microscopy images of an embryo simultaneously expressing YAP-miRFP670, CDX2-eGFP, mScarlet-I-SOX2, and H2B-miRFP720. Arrow: a cell that was classified as C^+S^+. A z-slice is shown. Scale bar: 10μm.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

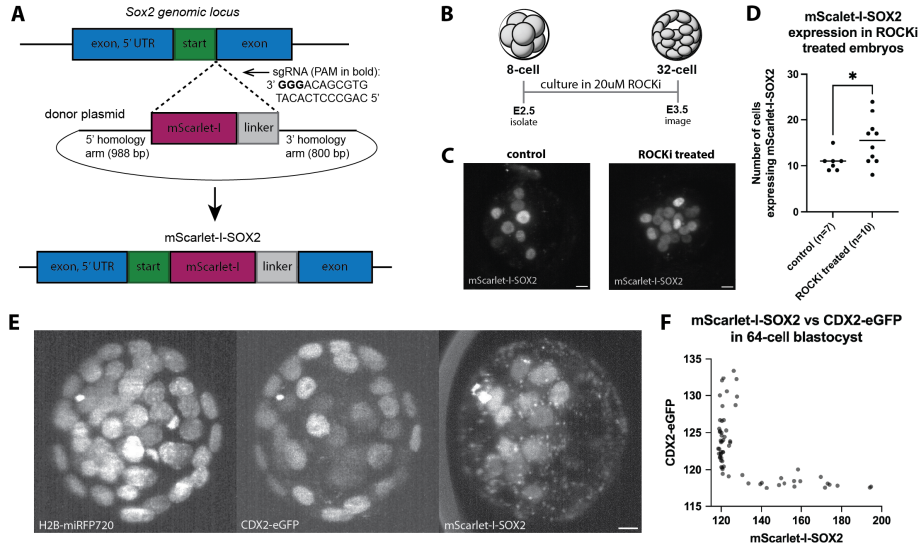


Figure S1: A) Targeting strategy for *mScarlet-I-Sox2* reporter mouse line. *mScarlet-I* plus a linker sequence was targeted to the N-terminus of the *Sox2* gene. B) Method for validation of *mScarlet-I-Sox2* reporter mouse line. 8-cell (E2.5) embryos were isolated and treated with 20μM ROCKi until the 32-cell stage (E3.5). At E3.5, ROCKi-treated embryos were live imaged for mScarlet-I-SOX2 alongside controls. C) Sample images of control vs. ROCKi-treated embryos, representative of $N = 7$ control, $N = 10$ ROCKi-treated embryos. Scale bar: 10μm. D) Quantification of the number of mScarlet-I-SOX2 expressing cells in control embryos versus embryos treated with 20μM ROCKi. In agreement with expected SOX2 behavior, embryos treated with ROCKi displayed significantly higher numbers of SOX2-expressing cells. Each point represents one embryo (Student's t-test, $p = 0.0432$). E) Live images of a 64-cell stage embryo expressing H2B-miRFP670 (nuclei), CDX2-eGFP, and mScarlet-I-SOX2. Scale bar: 10μm. F) Quantification of cells' expression levels of mScarlet-I-SOX2 versus CDX2-eGFP for the 64-cell stage blastocyst shown in panel (E). Each point represents one cell.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

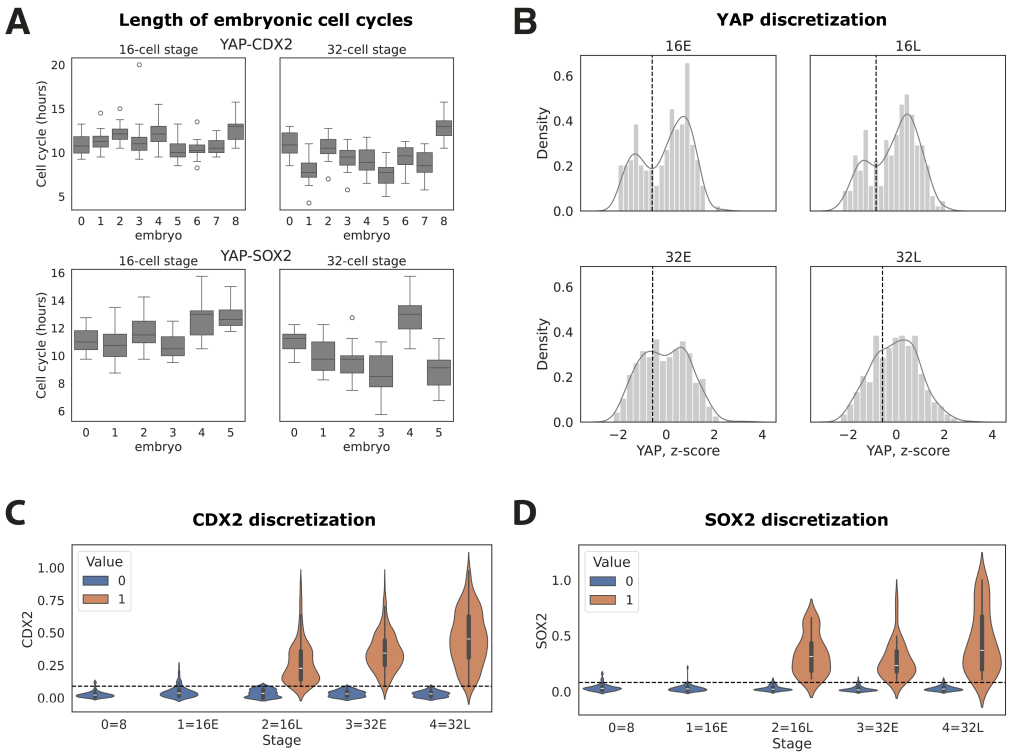
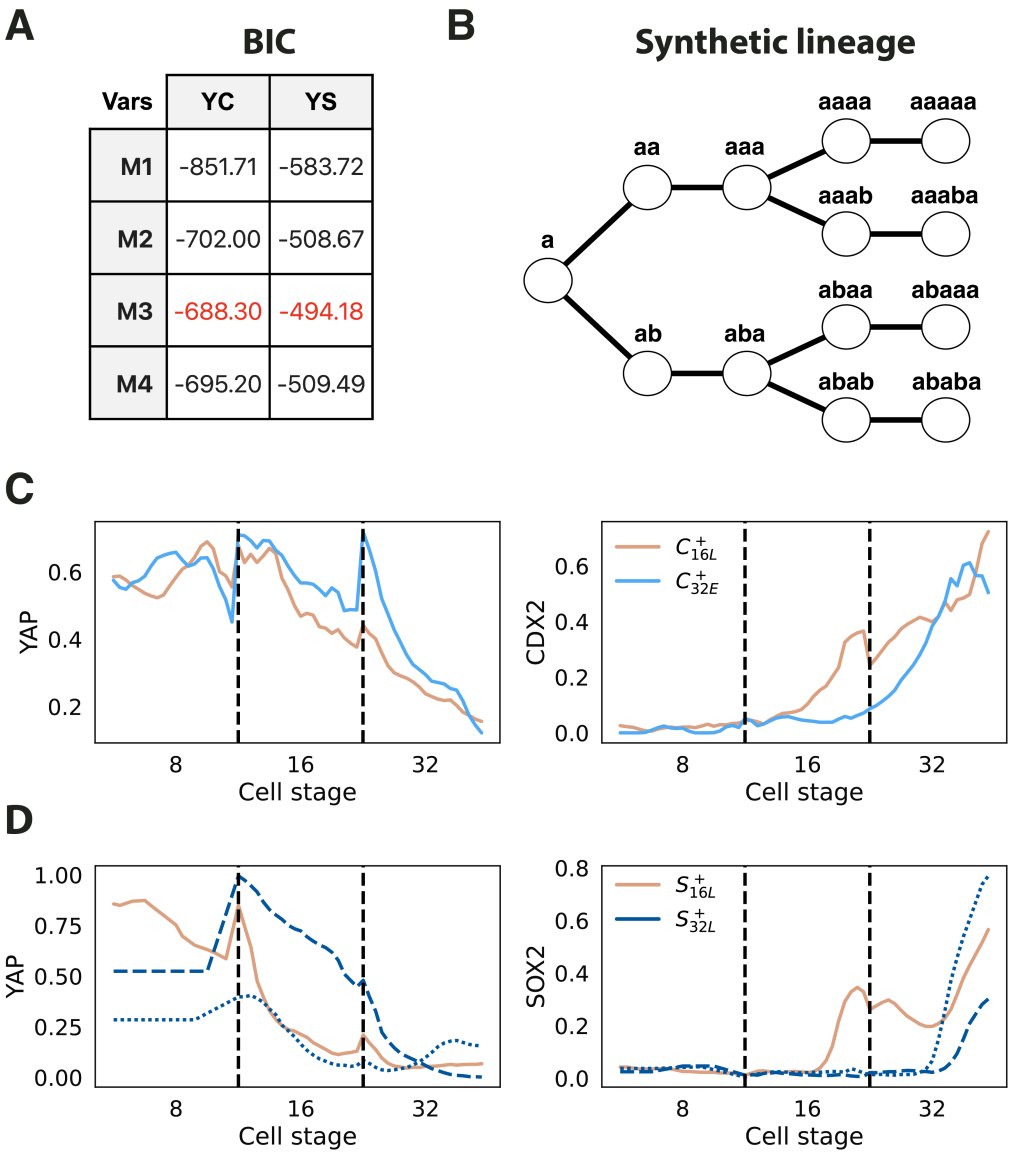


Figure S2: Data processing and discretization A) Boxplots summarizing the distributions of the cell cycle lengths in the embryos used in this study. x-axis: embryo, y-axis: cell cycle length, units are hours. Top row: YAP-CDX2 data, bottom row: YAP-SOX2 data. B) Histograms showing the distribution of summarized YAP values (Methods) for every discretized stage. Black vertical dashed line: stage-dependent threshold. C) Violinplots showing the distribution of summarized CDX2 values (Methods) for every discretized stage. Color: value after binarization. Black horizontal dashed line: stage-independent threshold. D) Same as C) for SOX2.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Vars	YC	YS
M1	-851.71	-583.72
M2	-702.00	-508.67
M3	-688.30	-494.18
M4	-695.20	-509.49

Figure S3: Bayesian model selection and inference details. A) Structure scores (BIC, see Methods) for the 4 candidate models from Fig. 3A. Rows: model, Columns: dataset; YC denotes the YAP-CDX2 dataset, YS denotes the YAP-SOX2 dataset. M3 model achieves the best (highest) structure score and was therefore selected for modeling both datasets. B) A schematic showing an encoding for every node in the synthetic lineages that were simulated from Bayesian networks. The lineages incorporate 2 rounds of divisions over 5 stages. C) Two representative examples of YAP-CDX2 traces demonstrating variable timing of CDX2 induction. Beige: a branch from C^{+}_{16L}, light blue: a branch from C^{+}_{32L}. Both branches were classified as nuclear YAP for all stages. D) Three representative examples of YAP-SOX2 traces demonstrating variable timing of SOX2 induction and variable corresponding YAP behaviors. Beige: a branch from S^{+}_{16L}, YAP is lost at 8/16 division for this branch. Dark blue: two branches from S^{+}_{32L}; dotted line: a branch with YAP lost at 8/16 division, dashed line: another branch with YAP lost at 16/32 division. x-axis: warped time, vertical black dashed line marks divisions.

bioRxiv preprint doi: https://doi.org/10.1101/2025.02.24.639895; this version posted February 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

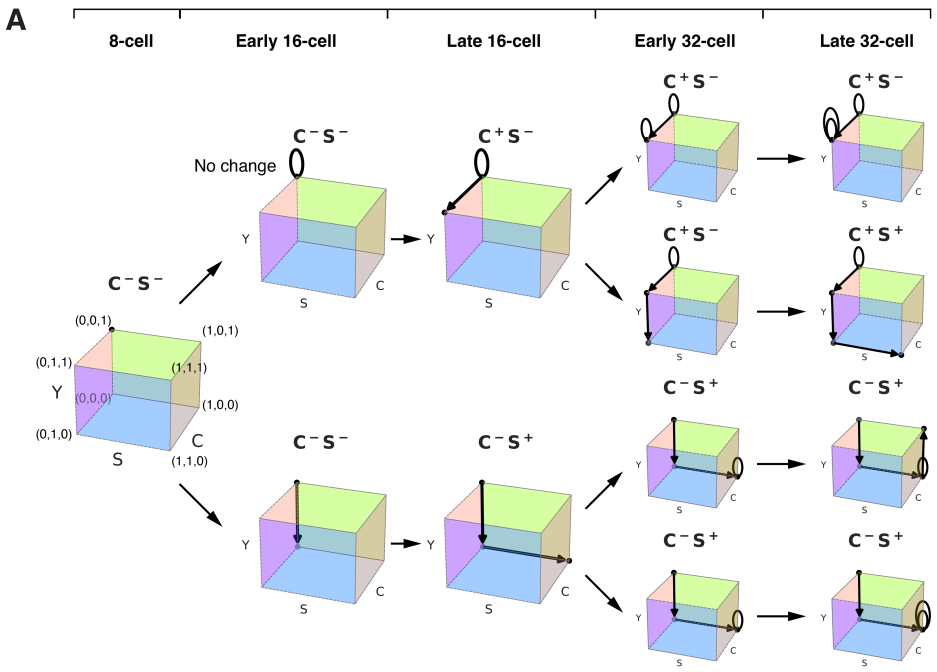


Figure S4: Data fusion with Bayesian networks. A) Expression of (S, C, y) in one lineage tree sampled from a fused model. Initially one cell at $(0, 0, 1)$ is sampled. At every step one or two cells are sampled using CPDs of the fused model. As a result, the trajectory of a cell is a random walk on the $\{0, 1\}^3$ cube, with two branching points resulting in 4 cells at 32L stage. Steps of the walk are shown with arrows. If the sampled value coincides with the previous one, a loop is shown. **(B)** Every simulated embryo is summarized as a table showing numbers of cells in C^+S^+, C^+S^-, C^-S^+, and C^-S^- classes over stages.