



Supplementary Materials for **Contact area–dependent cell communication and the morphological invariance of ascidian embryogenesis**

Léo Guignard*, Ulla-Maj Fiúza*, Bruno Leggio, Julien Laussu, Emmanuel Faure,
Gaël Michelin, Kilian Biasuz, Lars Hufnagel†, Grégoire Malandain†‡,
Christophe Godin†‡, Patrick Lemaire†‡

*These authors contributed equally to this work.

†Corresponding author. Email: hufnagel@embl.de (L.H.); gregoire.malandain@inria.fr (G.Ma.);
christophe.godin@inria.fr (C.G.); patrick.lemaire@crbm.cnrs.fr (P.L.)

‡These authors contributed equally to this work.

Published 10 July 2020, *Science* **369**, eaar5663 (2020)

DOI: 10.1126/science.aar5663

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S67
Tables S1 to S13
Captions for Movies S1 to S4
Captions for Data S1 to S18
References

Other Supplementary Material for this manuscript includes the following:

(available at science.sciencemag.org/content/369/6500/eaar5663/suppl/DC1)

MDAR Reproducibility Checklist (.pdf)
Movies S1 to S4 (.mov)
Data S1 to S18 (hosted on figshare—URLs are provided in this PDF)

Contents

1	Imaging of <i>Phallusia mammillata</i> embryos (Supplement to Figure 1)	7
1.1	Animals and embryological manipulations	7
1.2	Constructs and synthetic mRNA	7
1.3	Imaging of <i>Phallusia</i> embryos	8
2	Multi-angle fusion of 3D intensity image stacks (Supplement to Figure 1)	8
2.1	Definitions	8
2.2	Fusion	9
2.3	Additional module: slit line gain correction of acquired images	11
3	Assessment of the efficiency of MARS-ALT-based pipeline for sequential segmentation and tracking	12
3.1	Definitions	12
3.2	MARS segmentation	13
3.3	Cell tracking	14
3.4	Main limitations of MARS-ALT	16
4	One-pass ASTEC pipeline for simultaneous segmentation and tracking (Supplement to Figure 1)	16
4.1	Segmentation propagation pipeline	16
4.2	Initialization	16
4.3	Forward iteration overview	17
4.4	Segmentation projection	18
4.5	Cell division detection	19
4.6	Segmentation consistency checking	21
4.7	Post-processing	23
4.8	Additional module: region-based membrane enhancement	25
4.8.1	Tensor voting based membrane enhancement	25
4.8.2	Intensity 1-byte conversion	26
4.9	ASTEC parametrization	27
4.10	Time of execution	28
5	Cell naming and extraction of geometrical data	28
5.1	Naming of cells and fate allocation	28
5.2	Organization and extraction of geometrical data	31
6	Manual assessment and curation of ASTEC-segmented embryos after post-correction	32
6.1	Assessment of the quality of the automated segmentation and tracking of ASTEC-Pm1	32
6.2	Comparison with RACE, a state-of-the-art segmentation algorithm for cell membrane segmentation and tracking	33
7	Analysis of the stereotypy of <i>Phallusia mammillata</i> development (Supplement to Figure 2)	35
7.1	Temporal stereotypy	36
7.2	Spatial stereotypy	39
7.2.1	Cell volume comparisons	40
7.2.2	Analysis of relative positional variability	41

7.2.3	Cell neighbourhood comparisons	42
7.2.4	Comparison of areas of cell-cell contact	43
7.2.5	Comparative analysis of the degree of stereotypy in <i>P. mammillata</i> and <i>C. elegans</i>	45
7.2.6	Comparison between distributions	45
8	Analysis of fate restriction events during ascidian development (Supplement to Figure 3)	45
8.1	Automatic detection of cell fate restriction events	45
8.1.1	Detection of asymmetric cell divisions	45
8.1.2	Detection of geometrically unequal cell division and differential cell cycle duration	48
8.1.3	Clone shape analysis	49
9	Description of the differential sister cell induction model (Supplement to Figures 4 and 5A)	49
9.1	Model summary	49
9.2	An atlas of spatially-controlled signaling pathways during ascidian early embryogenesis	53
9.2.1	Signaling pathways with spatially-restricted gene expression profiles	54
9.2.2	Estimation of spatio-temporal protein expression	54
9.3	Principle of the cell contact areas model	55
9.4	Formalization of the signaling models	56
9.4.1	Computation of the positive signaling impact of a signaling ligand on an individual cell	57
9.4.2	Number of receptors complexed at a cell membrane	58
9.4.3	Number of effectors activated at a cell membrane	59
9.5	Computation of the combined signaling impacts of a ligand and its secreted antagonist on an individual cell	60
9.6	The case of mutually antagonistic FGF/Ephrin signaling	61
9.7	Deciding whether an induction takes place	63
9.7.1	Definitions	63
9.7.2	Algorithm to estimate of ϵ_{max} and ϵ_{min} for each pathway	64
9.7.3	Deciding on differential induction of sister cells.	65
9.8	Training the model	65
9.8.1	A census of known ascidian inductions	65
9.8.2	Exploration of parameter space	66
9.9	Description of the model results in wild-type conditions	67
9.10	Impact on the model output of perturbations in ligand expression profiles and embryo geometry	80
9.10.1	Long range ligands and inhibitors action	80
9.10.2	Comparison of model results obtained with membrane-based segmentation and Voronoï tessellation	80
9.10.3	Randomization of ligand gene expression pattern	81
9.10.4	Constant number of receptors at cell interfaces	81
9.10.5	Ephrin knock-out	82
10	Biological validation of the model output using live single cells measurements of ERK activity at the 64-cell stage (Supplement to Figure panels 5B)	82
10.1	KTR signal analysis	84
10.1.1	Processing of two-channel PH-tdTomato/ERK-KTR-Clover image datasets	84
10.1.2	Quantification strategy for the ERK-KTR signal	84

10.1.3	Time registration of ERK-KTR signals in different embryos	85
10.1.4	The subcellular localization of ERK-KTR correlates with ERK activity in ascidian embryos	86
10.2	Comparison of ERK-KTR signal to model predictions	87
11	Analysis of the effect of changes in the cell-cell contact areas (Supplementary information to Figure 5B-D)	88
11.1	Production, naming and analysis of half embryos	88
11.2	Response of the model to <i>in silico</i> changes in areas of cell-cell contacts	89
11.3	Comparison of the effect of changing areas of contact vs ligand expression levels	90
12	Author contributions.	92
13	Tables	94
14	Access to datasets and software.	108
14.1	Software and tutorials	108
14.2	Movies	108
14.3	Datasets	109

List of Figures

S1	Comparison of rigid and affine registration of intensity images acquired from two angles at the same time point	10
S2	Complementary contribution of individual views to the fused image	11
S3	Visualisation of a fused embryo	11
S4	Slit line gain correction	12
S5	h -minima operator	13
S6	Impact of h values on results of MARS segmentation on the <i>Phallusia</i> dataset at time $t = 152$, when the embryo counts 218 cells	14
S7	Impact of the segmentation errors types on cell tracking	15
S8	Impact of h values on lineage tracking with the MARS-ALT two-pass algorithm on the ASTEC-Pm1 dataset	16
S9	ASTEC iteration	17
S10	Illustration of the naming used for the registrations	18
S11	Number of seeds found for a cell \tilde{c} for different values of the parameter h for seed detection	19
S12	Distribution of h values selected during the ASTEC segmentation and tracking of the ASTEC-Pm1 dataset	20
S13	The ASTEC post-correction pipeline	23
S14	Membrane enhancement	26
S15	ASTEC-Pm1 cell lineage trees with associated fate map	30
S16	Position of individual tissue precursors at the early tailbud stage in ASTEC-Pm1	31
S17	Cell smoothing	32
S18	Quantification of the quality of the shape of ASTEC-segmented cells in ASTEC-Pm1	33
S19	Comparison ASTEC - RACE	35
S20	Evolution of compactness during mitosis	36
S21	Comparison of ASTEC cell lineage trees from bilateral A7.4 blastomeres	37

S22	Cell cycle duration of matching cells	38
S23	Principle of tree-edit distance computation between lineages	38
S24	Analysis of cell volumes across cell division in ASTEC-Pm1	40
S25	Volume of matching cells	41
S26	Cell position variability	42
S27	Duration of contacts	44
S28	Division distances between lineages	46
S29	Hierarchical clustering of individual cells at the 64-cell stage, according to the tree-edit distance between the lineage trees they seed	47
S30	Sister cell volumes or cell cycle durations ratios in ASTEC-Pm1	49
S31	Shapes of 4-cell clones	50
S32	Pattern of expression of RNA and protein for the major signaling molecules up to the early gastrula stage	53
S33	Types of signaling molecules involved and effect on inductions	57
S34	Example of the relationship between $\frac{(k_e+k_d)}{k_e} \cdot \frac{E_i^*}{E_i^T}$ and the area of contact A_i between emitting and responding cells	61
S35	Defining a minimal level of signaling necessary for inductions to occur	64
S36	Signaling thresholds	66
S37	Result for the model run under wild type conditions	68
S38	Result for the model run under wild type conditions for ASTEC-Pm5	69
S39	Result for the model run under wild type conditions for ASTEC-Pm8	70
S40	\bar{E}^* distribution, ERK, stage 6	71
S41	\bar{E}^* distribution, ERK, stage 8	71
S42	\bar{E}^* distribution, ERK, stage 10	72
S43	\bar{E}^* distribution, ERK, stage 11	72
S44	\bar{E}^* distribution, Bmp, stage 6	73
S45	\bar{E}^* distribution, Bmp, stage 8	73
S46	\bar{E}^* distribution, Bmp, stage 10	74
S47	\bar{E}^* distribution, Bmp, stage 11	74
S48	\bar{E}^* distribution, Nodal, stage 6	75
S49	\bar{E}^* distribution, Nodal, stage 8	75
S50	\bar{E}^* distribution, Nodal, stage 10	76
S51	\bar{E}^* distribution, Notch, stage 8	76
S52	\bar{E}^* distribution, Notch, stage 10	77
S53	\bar{E}^* distribution, Notch, stage 11	77
S54	\bar{E}^* distribution, Wnt, stage 6	78
S55	\bar{E}^* distribution, Wnt, stage 8	78
S56	\bar{E}^* distribution, Wnt, stage 10	79
S57	\bar{E}^* distribution, Wnt, stage 11	79
S58	Number of optimal parameter values for the model run with first or second degree neighbors action range for secreted signalling proteins	81
S59	Results of the model run when Ephrin receptor signaling is computationally inhibited	83
S60	Examples of dynamic ERK-KTR nucleo-cytoplasmic shuttling in two example cells	84
S61	Schematic flow of the iterative process of time registration of KTR signal.	86
S62	Average time-registered inverse normalized ERK-KTR value for wild-type embryos	87

S63	Time-registered inverse normalized ERK-KTR value for embryos treated with the U0126 MEK inhibitor from the early 16-cell stage.	88
S64	Effect of halving embryos on the neighborhood, ERK signalling intensity and cell lineage of the a6.5 and a6.6 blastomeres.	89
S65	ERK activation in the half embryo ASTEC-half-Pm1	90
S66	Sensitivity of induced states to changes in the surface of contact to signalling cells. . . .	91
S67	Comparison of the effect of altering the geometry of cell interfaces or ligand expression levels on the ERK pathway induction status of cells.	91

List of Tables

S1	<i>Phallusia mammillata</i> datasets	95
S2	Automatic post-corrections	96
S3	Manual corrections and assessment of final segmentation and tracking	97
S4	Ground truth for known asymmetric cell divisions	98
S5	Ground truth for known symmetric cell divisions II	99
S6	Asymmetric division found by the classifier for known and candidate asymmetric cell divisions (fate restriction events).	100
S7	Volume ratio between sister cells	101
S8	Volume ratio between sister cells II	102
S9	Cell cycle length ratio between sister cells	103
S10	Cell cycle length ratio between sister cells II	104
S11	List of experimentally-determined differential sister inductions and mother polarizations leading to differential fates acquisition by sister cells	105
S12	Model parameters comprehensive list and their optimal values	106
S13	Model predictions and parametrization	107

List of Movies

S1	3D Rendering of intensity	108
S2	3D Rendering of intensity for 10 embryos	108
S3	ASTEC segmentation, clonal view	108
S4	ASTEC segmentation, Fate map view, 10 embryos	108

List of Datasets

S1	Astec-Pm1, Wild type	109
S2	Astec-Pm2 Wild type	110
S3	Astec-Pm3, Wild type	110
S4	Astec-Pm4, Wild type	110
S5	Astec-Pm5, Wild type	110
S6	Astec-Pm6, Wild type	110
S7	Astec-Pm7, Wild type	110
S8	Astec-Pm8, Wild type	110
S9	Astec-Pm9, Wild type	111

S10 Astec-Pm10, Wild type	111
S11 Astec-Pm11, Wild type	111
S12 Astec-Pm12, Wild type	111
S13 Astec-Pm13, Wild type	111
S14 Astec-Pm14, Wild type	111
S15 Astec-Pm15, Wild type	111
S16 Astec-U0126-Pm1, U0126 treated	111
S17 Astec-U0126-Pm2, U0126 treated	112
S18 Astec-half-Pm1, half embryo	112

1 Imaging of *Phallusia mammillata* embryos (Supplement to Figure 1)

1.1 Animals and embryological manipulations

Phallusia mammillata were provided by the Centre National de Ressources Biologiques Marines in Roscoff (France). Embryos were obtained by *in vitro* fertilization and dechorionated as in (41).

Phallusia mammillata unfertilized eggs were microinjected under a stereoscope, with approximately 45pg of mRNA of construct as in (42).

The ERK signalling pathway was blocked with the pharmaceutical MEK inhibitor U0126 (abcam : ab120241) (43). U0126 was diluted in artificial sea water, used at concentrations of to 2 μ M or 6 μ M as described in (42) and applied from the early 16-cell stage to the end of the imaging period.

To experimentally manipulate the surface of contact between cells, we bisected microinjected embryos along their plane of bilateral symmetry at the 2-cell stage with a thin custom-made glass tool. Half embryos were transferred at the equivalent of the 16-cell stage to the MuViSPIM microscope for imaging (Table S1).

1.2 Constructs and synthetic mRNA

Construction of the pGEM-PH-citrine plasmid

A pGEM-HE vector (44) was modified by introducing a new multiple cloning site containing Mlu/XhoI/SalI restriction sites between the pGEM-HE BglII and SalI restriction sites, using synthetic oligonucleotides (Fwd: 5' – GATCTACGCGTAACCTCGAGAACG – 3'; Rev: 5' – TCGACGTTCTCGAGGTTACGCGTA – 3'). The sequence encoding the yellow fluorescent protein mCitrine was amplified by PCR to introduce MluI/XhoI and SalI/STOP codon flanking regions with synthetic primers (Fwd: 5' – CTAACGCGTAACCTCGAGATGGTGAGCAAGGGCGAGGAGCTG – 3'; Rev: 5' – TTGGTCGACTCACTGTACAGCTCGTCCATGCC – 3'). The PH domain of the human phospholipase C delta 1 (PLCD1) coding sequence from pRN3-PH-GFP (45) was amplified by PCR to introduce MluI/kozak and XhoI flanking regions with synthetic primers (Fwd: 5' – AATACGCGTCAGAAAAATGGACTCGGGC-CGGGACTTCC – 3'; Rev: 5' – AACCTCGAGGATGTTGAGCTCCTTCAGGAAG – 3'). These two PCR inserts were sequentially introduced into the modified pGEM-HE vector. The construct is available in AddGene (id: 131406).

Construction of the pSpe3-ERK-KTR-clover plasmid

The pSpe3-ERK-KTR-clover plasmid pSpe3-ERK-KTRclover was built by recombining the gateway construct pENTR-ERK-KTRclover (a gift from Markus Covert; Addgene plasmid #59138; (46)) with the pDEST/Spe3-Rfa mRNA synthesis vector, as described in (47).

In vitro mRNA synthesis

mRNAs for microinjection were synthesized using the mMessage mMachine T3 and T7 Ambion transcription kits using as templates the pRN3-PH-GFP construct (a kind gift from Alex McDougall, Obser-

vatoire Océanologique de Villefranche-sur-Mer, France; (45)), the pRN3-PH-tdtomato (a kind gift from Yasuo Hitoyoshi, Observatoire Océanologique de Villefranche-sur-Mer, France), pSpe3-ERK-KTR-clover and pGEM-PH-citrine.

1.3 Imaging of *Phallusia* embryos

Embryos were imaged in artificial seawater using either a custom multiview lightsheet microscope at EMBL (MuViSPIM; (48)), or a commercial derivative in Montpellier (MuViSPIM, Luxendo). Embryos showing the canonical pattern of cell cleavages at the 16- to 64-cell stage were selected for imaging, deposited by gravity without embedding at the bottom of a 0.8% GelRite (SIGMA, G1910) or 1% Low Melt Agarose (Carl Roth, 6351.1) cone-shaped well and imaged at a constant temperature of $\sim 18^\circ\text{C}$ ($\sim 16^\circ\text{C}$ for ASTEC-Pm1) (Table S1).

PH-GFP and ERK-KTR-clover were excited with a 488 nm laser (LuxX 488-60, Omicron), PH-citrine with a 514 nm laser (OBIS 514-LX40mW, Coherent) and PH-tdtomato with a 561 nm laser, with simultaneous two-sided illumination. The emitted light was collected by two opposing detection arms positioned perpendicularly to the illumination plane, resulting in simultaneous dual acquisition. The movies ASTEC-Pm1 and ASTEC-Pm2 were imaged in a MuViSPIM set up where each detection arm was equipped with a 25x Nikon water dipping objective (NA 1.1) combined with a tube lens with a focal length of 300 mm leading to a 37.5 fold image magnification. Imaging of the other embryos was done in set ups where each detection arm was equipped with a 20x Olympus water dipping objective (NA 1.0) combined with a tube lens leading to a 33.3 fold image magnification. The filters used were the band-path BrightLine 525/30 filter (Semrock), the 528/38 filter (Semrock) and the 561LP filter. Filtered emitted light was collected by Hamamatsu V2 Flash 4 SCMOS cameras. An electronic confocal slit detection (eCSD) mode was used during image acquisition to minimize the capture of scattered light and improve contrast (implementation described in (13)). The electronic slit size was matched in accordance to the diameter of the illumination beams (60 pixels).

At each time point, two acquisitions were sequentially performed, the second one orthogonal to the first. For each embryo and each time point, this process generated four whole-embryo stacks, i.e. two sets of two matching 3D intensity image stacks (0, 90, 180, 270 degrees), comprising a selected volume size covering the imaged embryo. In case of double labeling experiment (PH-tdtomato and ERK-KTR-clover), two stacks, one for each label, were successively captured by each camera along each axis (membrane then ERK; 90 deg rotation; membrane then ERK). The embryos were imaged every two minutes. The ASTEC-Pm1 and ASTEC-Pm2 movies, were acquired with an image voxel size of $0.173 \times 0.173 \times 1 \mu\text{m}^3$. The remaining movies were acquired with an image voxel size of $0.195 \times 0.195 \times 1 \mu\text{m}^3$. The intensity values at each voxel were encoded in 16 bits.

2 Multi-angle fusion of 3D intensity image stacks (Supplement to Figure 1)

2.1 Definitions

Using the above protocol, 3-dimensional (3D) images were acquired at different time points t and along different angles $a \in \{1, 2, 3, 4\}$. Each image, denoted by I_t^a , is a 3D array of intensity values:

$I_t^a: \Omega \subset \mathbb{R}^3 \rightarrow \mathcal{I} \subset \mathbb{N}$. I_t^a associates each 3D point in the image definition space with a fluorescence intensity value $\mathcal{I} \in [0, 2^{16} - 1]$.

At each time point t , the image obtained from the fusion of the images I_t^a observed from the different angles $a \in \{1, 2, 3, 4\}$, is denoted by $I_t: \Omega \subset \mathbb{R}^3 \rightarrow \mathcal{I} \subset \mathbb{N}$ and $\{I_t\}_{t \in \{t_1, \dots, t_f\}}$ is a temporal sequence of 3D fused intensity images indexed by the consecutive observation time points $\{t_1, \dots, t_f\}$. When necessary, we will denote $I_{t,\sigma}$ the fused 3D intensity image at time t after applying a Gaussian smoothing with parameter σ .

2.2 Fusion

Individual image stacks from each camera were first automatically cropped and resampled to a resolution of $0.3 \times 0.3 \times 0.3 \mu\text{m}^3$. The cropping was done by first computing the maximum intensity projection (MIP) of the image along the z axis. Then the MIP image was binarized using its mean value as a threshold and only the largest connected components were kept. The bounding box of the resulting image was then enlarged by 40 voxels on each side to define the cropped image (in xy -plane). The resampling was done using tri-linear interpolation. The four images from a given time point ($\{I_t^a\}_{a \in \{1, \dots, 4\}}$) were then fused to build a 3D isotropic image I_t of the embryo. This fusion was done in four steps.

First, to compensate for the reduction in image quality caused by the distance traveled by emitted photons through the embryo before reaching the objective, a weight function $w^a(v)$ on image points $v = [x, y, z]^T$ was computed for each image I_t^a :

$$w^a(v) = 0.1 + e^{-\lambda^a d^a(z)}, \quad (2.1)$$

where $d^a(z)$ is the distance traveled through regions of the 3D image with an intensity higher than the Otsu threshold (49), between the illuminated plane and the boundary of the cropped image on the path to the camera and $\lambda^a = 5/l^a$, l^a being the maximum of the function d^a over the whole image. The factor 5 was chosen empirically.

Second, all four images were registered onto the referential of the first one I_t^1 , using a rigid transformation. This rigid registration was not sufficient to correct the minor distortions originating from slight misalignment of the cameras (Fig. S1, central panels).

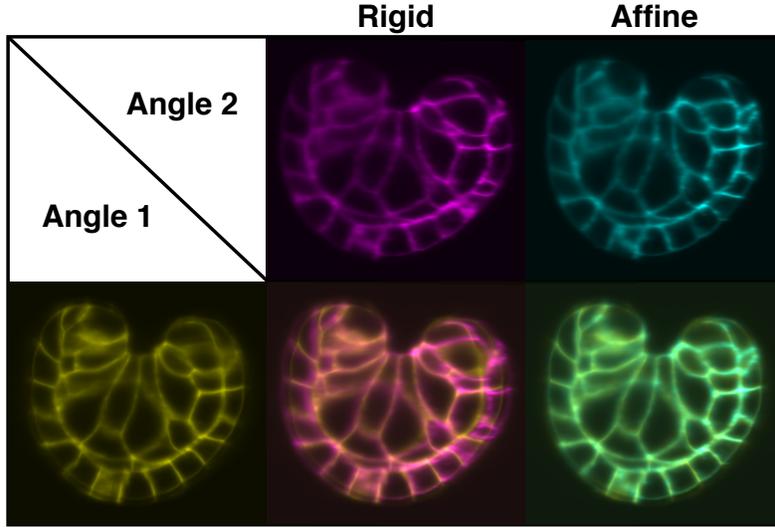


Figure S1: Comparison of rigid and affine registration of intensity images acquired from two angles at the same time point. The figure shows matching cross-sections through a gastrula embryo (time $t = 152$ min fixed), extracted from the <https://translate.google.fr/?hl=fr> image stacks from two different angles. The floating image (angle 2, first row) is registered onto the reference image (angle 1, yellow) with either a rigid (magenta) or an affine (blue) transformation. The bottom row shows the overlay of the referential image with the rigid or affine registration of the floating image.

Third, these distortions were corrected by computing the three affine transformations $\mathcal{T}_{a \leftarrow 1}$ that map the frame of I_t^1 onto I_t^a frames, and thus allow to resample I_t^a images into the I_t^1 frame (block-matching algorithm described in (50, 51)), Fig. S1 right panels.

Finally, the fused isotropic image I_t was generated in the I_t^1 frame, in which the intensity of each point v is the weighted average of the intensity values of the corresponding voxel in the four registered images:

$$I_t(v) = \sum_a \frac{w^a \circ \mathcal{T}_{a \leftarrow 1}(v)}{\sum_a w^a \circ \mathcal{T}_{a \leftarrow 1}(v)} I_t^a \circ \mathcal{T}_{a \leftarrow 1}(v) \quad (2.2)$$

Fig. S2 illustrates the quality improvement obtained by the fusion process, Fig. S3 illustrates fused images early and late in the development.

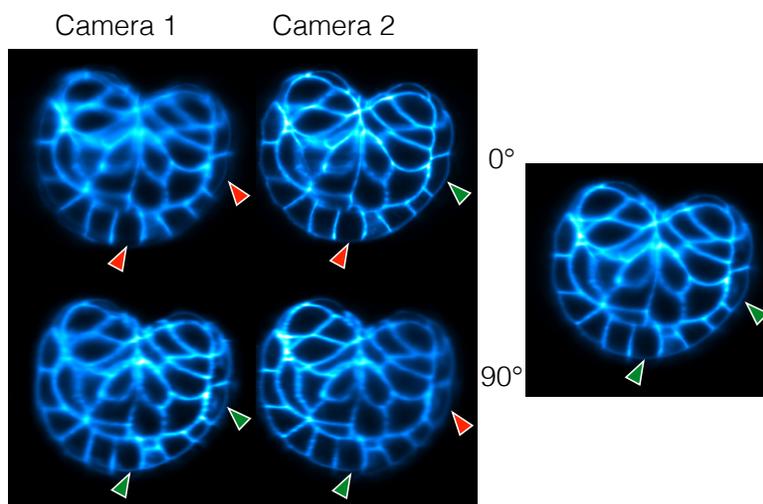


Figure S2: Complementary contribution of individual views to the fused image Left: matching optical sections through the same embryo from 4 different angles of views. Right: Corresponding fused image. Arrowheads point to matching membranes of interest. Green arrowheads: high quality signal. Red arrowheads: faint or absent signal.

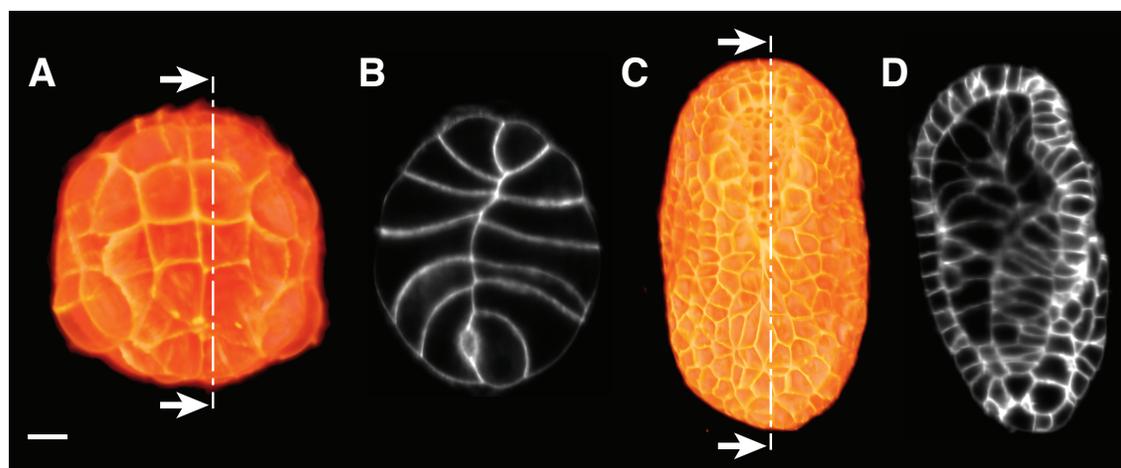


Figure S3: Visualisation of a fused embryo A) Vegetal view of the embryo at the 64-cell stage (stage 8, time $t = 1$). B) Sagittal section along the plane shown on A. C) Dorsal view of the embryo at the late neurula stage (Stage 16, time $t = 180$). D) Sagittal section along the plane shown on D. Anterior is to the top. Scale bar: $20\mu\text{m}$.

2.3 Additional module: slit line gain correction of acquired images

The Hamamatsu Flash 4 V2 cameras used in the MuViSPIMs used in this study have a two component detector resulting in brighter XZ sections at the boundary between the two components in images acquired in slit mode (see Fig. S4, second panel from the left) which may be interpreted as membrane during segmentation. We designed a procedure to correct this artefact before the fusion.

1. The 50 % Y -lines with the smallest average values are selected, yielding a set of Y -lines totally included in the background (without membrane signal), supposed to be of uniform values.

2. For each selected line, the 10 % points with the largest difference with the average are detected. Those points are likely to belong to the XZ -sections affected by the artefact.
3. XZ -sections to be corrected are then selected based on their occurrence in the discarded values (i.e. they appear in more than 20 % of the above detected points).
4. A linear correction is then computed for the selected XZ -sections based on their comparison with the neighbouring unaffected XZ sections, which kept their original values.

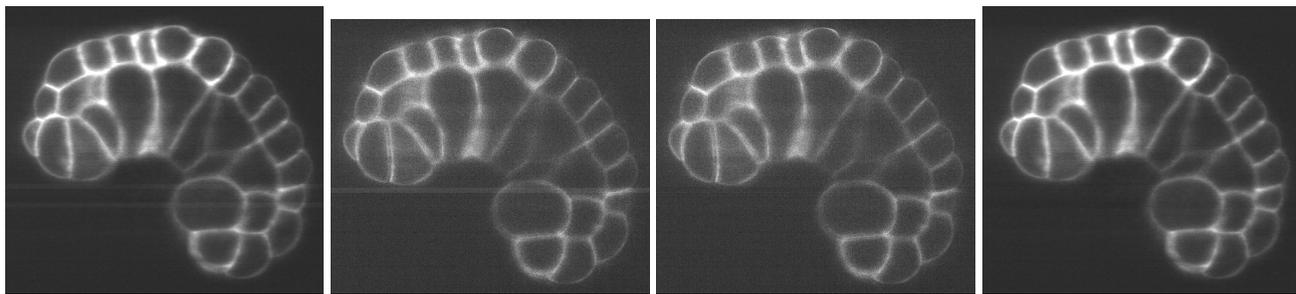


Figure S4: slit line gain correction From left to right: a XY -section of a fused image without correction; a XY -section of one individual stack from one camera before correction, some consecutive horizontal lines have a larger gain than the others, and may be interpreted as a spurious membrane (intensity windowing has been exaggerated for better visualization) - since these lines appear in the four acquired images, the fusion process may duplicate them in the fused image; the same image after correction; the fused image after correction of the 4 acquisitions.

3 Assessment of the efficiency of MARS-ALT-based pipeline for sequential segmentation and tracking

We previously developed a two-pass MARS-ALT pipeline in the context of confocal imaging of plant tissues (20). MARS-ALT proceeds in two passes. The first pass, MARS, produces a set of segmented images $\{S_t\}_{t \in \{t_1, \dots, t_f\}}$ from the original sequence of fused intensity images $\{I_t\}_{t \in \{t_1, \dots, t_f\}}$. In the second pass, correspondences between cell snapshots in consecutive segmented images are identified, leading to the construction of cell lineages. In this section, we discuss the results obtained with this pipeline on high-throughput MuViSPIM imaging of *Phallusia* embryo development. A precise understanding of the causes of failure of MARS-ALT was important for the development of a novel single-pass method, ASTEC described in the following section.

3.1 Definitions

A segmented image is an image $S_t: \Omega \subset \mathbb{R}^3 \rightarrow C_t \subset \mathbb{N}$ where C_t is a finite set of labels identifying in a unique manner the region of each segmented cell in the embryo at time t . By convention, label 1 is always allocated to the region of the image outside of the embryo. Label 0 is reserved for the temporary identification of untreated voxels in the post-process operations, such as the watershed or seed detection (see below). $\{S_t\}_{t \in \{t_1, \dots, t_f\}}$ defines a temporal sequence of 3D segmented images.

To ensure consistency of the labels throughout a given sequence $\{S_t\}_{t \in \{t_1, \dots, t_f\}}$, each cell snapshot - the image of a cell at a given time point - has a unique identifier throughout space and time:

$$\forall i, j \in \{t_1, \dots, t_f\}^2, i \neq j \Rightarrow C_i \cap C_j = \emptyset \quad (3.1)$$

3.2 MARS segmentation

To segment original images, MARS uses a 3D watershed algorithm with prior automatic seed detection, (20). Seeds are identified using the h -minima operator **Hmin** (52) to find the set of local connected intensity minima regions that are separated by a minimum intensity difference of h across a smoothed fused image I_{t, σ_1} (Fig. S5). Voxels that do not belong to a seed are labeled 0 and the exterior is considered as a cell that has to be seeded as the other cells. For a given h , this step results in a pre-segmented image with seed labels, denoted by $Seeds_{t, \sigma_1}^h$:

$$Seeds_{t, \sigma_1}^h = \mathbf{Hmin}(I_{t, \sigma_1}, h). \quad (3.2)$$

Pre-segmented seeded images $Seeds_{t, \sigma_1}^h$ are then used to segment a less smoothed image I_{t, σ_2} , $\sigma_2 < \sigma_1$, with a 3D watershed algorithm, **WS**, to produce a segmented image S_t^h :

$$S_t^h = \mathbf{WS}(Seeds_{t, \sigma_1}^h, I_{t, \sigma_2}). \quad (3.3)$$

The values of the σ parameters were chosen empirically $\sigma_1 = 0.6 \mu\text{m}$ (strong smoothing for seed detection) and $\sigma_2 = 0.15 \mu\text{m}$ (weak smoothing for membrane detection).

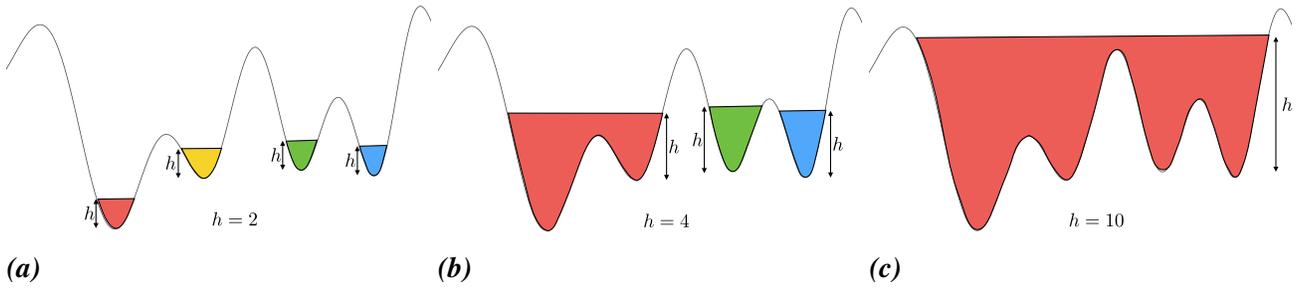


Figure S5: h -minima operator. The figure shows how changes in h values impact the number of detected seeds. a) Local minima detection, $h = 2$: 4 seeds are detected; b) Local minima detection, $h = 8$: 3 seeds are detected; c) Local minima detection, $h = 11$: only 1 seed is detected.

In MARS, the quality of segmentations is particularly sensitive to the seed detection step of the algorithm (20). Four types of seed detection errors can occur (Fig. S7).

1. Detecting two or more seeds within the boundaries of a single cell snapshot. This oversegmentation error splits a cell into two or more, and is denoted as \mathbf{E}_s error type.
2. Detecting more than one seed outside of the embryo. This error results in the addition of an artefactual extra cell to the embryo (\mathbf{E}_a).
3. Failing to detect a seed within the boundaries of a cell. This error can result in two different types of under-segmentation outcomes:
 - (a) the region of the missed cell is allocated to the exterior and the corresponding cell snapshot is missing from the embryo (\mathbf{E}_{mi}),
 - (b) the region of the missed cell is allocated to a neighbouring cell snapshot, merging these two cells together (\mathbf{E}_{me}).

A fifth class of segmentation errors is produced during the watershed step as a result of inaccurate seeding, when voxels are not allocated to the correct cell snapshot. This error results in cell shape defects (\mathbf{E}_{sh}).

MARS uses a unique h value for all cells in the sequence. To analyze the impact of h on seed detection errors, MARS was run on the fused 3D intensity image at time 152 of the ASTEC-Pm1 fused movie, with h values ranging from 4 to 18, by increments of 2 (Fig. S6). No single h value optimized all error types across the embryo. High h values reduced over-segmentation ($\mathbf{E}_s + \mathbf{E}_a$) but increased the number of undetected cells ($\mathbf{E}_{mi} + \mathbf{E}_{me}$), resulting in more fused cells. Low h values improved cell detection, but caused more frequent over-segmentation. While no h value gave a perfect segmentation, the process is however quite efficient, with less than 5% of segmentation errors for $h = 18$.

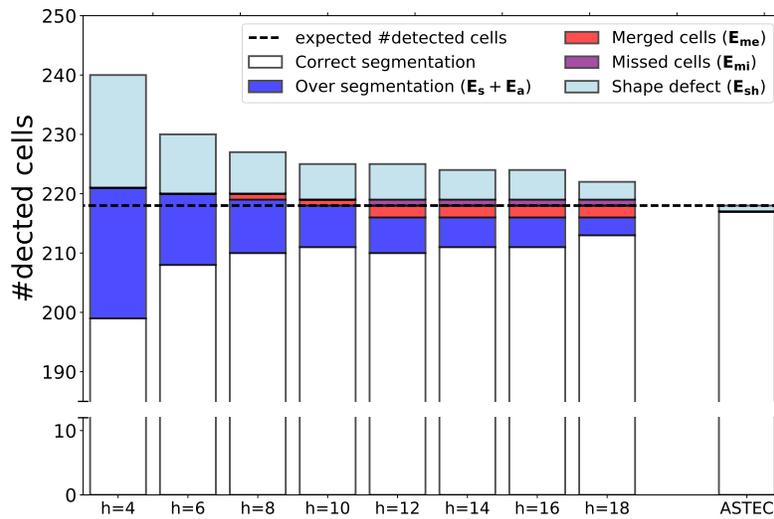


Figure S6: Impact of h values on results of MARS segmentation on the ASTEC-Pm1 dataset at time $t = 152$, when the embryo counts 218 cells. Note that no single h value optimizes all scores: smaller values avoid missed and merged cells, while high values reduce over-segmentations. For this plot, a cell is considered to have an error of type \mathbf{E}_{sh} if its dice score to its corresponding ground truth cell is less than .9.

3.3 Cell tracking

Given the high temporal resolution of our dataset, embryo geometries at consecutive time points were sufficiently similar to use a simpler version of the ALT tracking procedure, without impacting tracking quality. The modified version proceeds as follows.

Cell tracking proceeded by rigidly registering S_t on S_{t+1} and identifying in the segmented image S_t at time t the ancestor p of each cell region q in S_{t+1} . For this, we first computed the affine transformation mapping the segmented image S_t onto the segmented image S_{t+1} . This produces a registered image $S_{t+1 \leftarrow t}$. Then, for each cell region q in S_{t+1} , we computed its pairwise Dice scores \mathcal{D} with every cell region p in $S_{t+1 \leftarrow t}$ intersecting q : $\mathcal{D}(p, q) = \frac{2|p \cap q|}{|p| + |q|}$. This score identifies the cell region p in S_t that spatially best matches the region q . In the absence of any segmentation error, this procedure allocates

exactly one ancestor at time t to each cell region at time $t + 1$. In the case of a cell division, two regions at time $t + 1$ may match the same ancestor region at time t .

Depending on their type and co-occurrence in the image, segmentation errors had distinct effects on cell tracking (Fig. S7). Isolated E_s and E_a errors lead to artefactual cell divisions restricted to a single cell lineage. By contrast, E_{me} and E_{mi} errors may in addition artefactually connect neighbouring lineages. Note that all four error types lead to the appearance of artefactual cell divisions, and therefore the creation of artefactual branches in the inferred cell lineage.

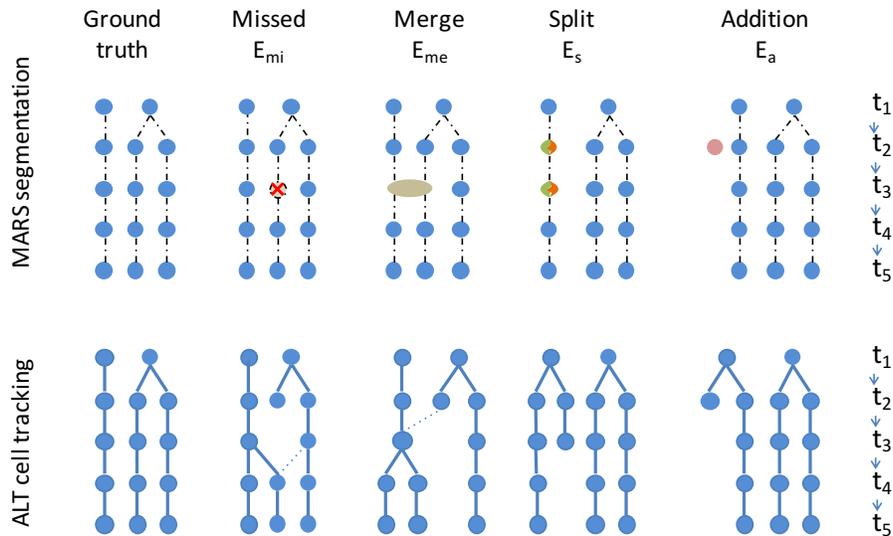


Figure S7: Impact of the segmentation error types on cell tracking. The top panels illustrate simple cases of the four major types of segmentation errors. Dashed lines show true lineage cell relationships. The bottom panels illustrate how each error is handled by the ALT tracking algorithm. Dashed lines represent potential alternative connections.

Analysis of the lineage trees obtained with a range of h values (Fig. S8) shows that a majority of cells are abnormally short-lived (cell cycle duration < 10 minutes), and that 36 to 44% of cell lineages were interrupted before the end of the movie, which is artefactual as no cell death is expected before metamorphosis (22, 53).

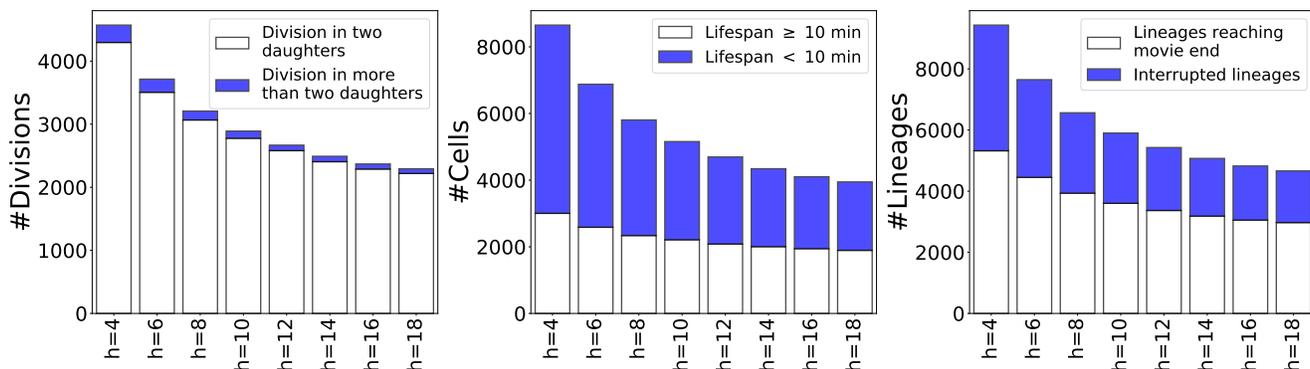


Figure S8: Impact of h values on lineage tracking with the MARS-ALT two-pass algorithm on the ASTEC-Pm1 dataset

3.4 Main limitations of MARS-ALT

The previous analyzes thus reveal that MARS-ALT is inadequate to segment high-throughput MuViSPIM datasets. Two major limitations were identified. First, using a uniform h value across all cell snapshots of the embryo leads to at least 5% of segmentation errors (Fig. S6). This small percentage of segmentation errors at each individual time point, however, led to the progressive accumulation of tracking errors, including the detection of precocious or artefactual cell divisions in most cells, and the premature interruption of 36 to 44% of cell lineages (Fig. S8).

The solution presented in the following section was designed to: 1) improve segmentation efficiency by adapting the chosen h value to the local environment of the cell; 2) improve cell tracking by propagating the segmentation between consecutive time points, rather than independently segmenting each time point.

4 One-pass ASTEC pipeline for simultaneous segmentation and tracking (Supplement to Figure 1)

4.1 Segmentation propagation pipeline

ASTEC (Adaptive Segmentation and Tracking of Embryonic Cells) is a novel single-pass algorithm, which simultaneously performs cell segmentation and tracking of the 4D sequence by propagating segmentations between consecutive time points of the image sequence $\{I_t\}_{t \in \{t_1, \dots, t_f\}}$, a strategy pioneered for nuclear labels by Amat and colleagues (54). The pipeline includes: 1) an **initialization phase**, during which a reference manually-curated segmentation of the first time point is generated; 2) an **iteration phase**, during which segmentations and tracking are propagated forward in time between consecutive time points up to the end of the sequence; 3) a **post-processing phase**, during which both cell lineages and segmentations are corrected. We detail thereafter these different steps.

4.2 Initialization

To initiate the process, a segmentation S_{t_1} of the image I_{t_1} is first computed using the MARS segmentation (20) algorithm and manually corrected. The same σ values σ_1 and σ_2 were used as in the two-pass

algorithm. To ease the post-correction phase, a small h parameter was chosen ($h = 4$) to favor over-segmentations and avoid undersegmentation during this initialization step. Oversegmented cells were manually fused using Fiji (55) to create a semi-automatic curated segmentation $S_{t_1}^*$ of I_{t_1} .

4.3 Forward iteration overview

Assuming by induction that a segmentation S_t^* has been obtained at time t , the algorithm propagates S_t^* onto S_{t+1}^* at $t + 1$ to build the segmentation at time $t + 1$. This is done in 3 steps (Fig. S9).

1. **Segmentation projection:** based on image registration between consecutive time points, the algorithm first estimates, for each cell $c \in C_t$ of S_t^* , the space potentially occupied by its progeny in I_{t+1} . This step produces a segmented image \check{S}_{t+1} . To each cell $c \in C_t$ corresponds a unique cell in \check{S}_{t+1} , as this step does not capture cell division events.
2. **Cell division detection:** to identify cells that divided between t and $t + 1$, this step performs a local seed detection within each cell of \check{S}_{t+1} to identify putative new membranes resulting from cytokinesis. This step produces the segmented image \hat{S}_{t+1} .
3. **Segmentation consistency checking:** the \hat{S}_{t+1} segmentation may be different from \check{S}_{t+1} either because of additional cells due to cell divisions or because of errors. These discrepancies between the two segmentations are detected and resolved. This last step produces S_{t+1}^* , the final segmentation.

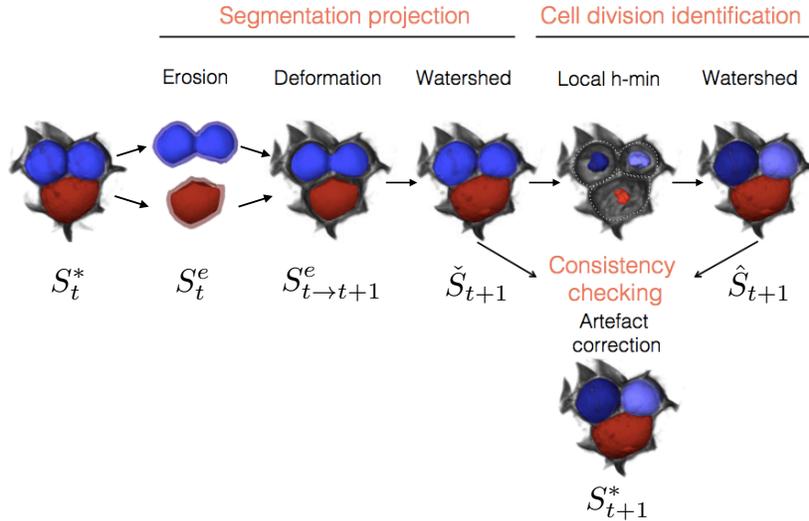


Figure S9: ASTEC iteration: starting from the segmented image at time t , ASTEC first erodes each cell (the blue cell is in mitosis, the red cell in interphase) from S_t^* to produce S_t^e . It then projects each eroded cell from S_t^e onto the referential of I_{t+1} to generate $S_{t \rightarrow t+1}^e$. \check{S}_{t+1} is the result of a local watershed using the eroded cells $S_{t \rightarrow t+1}^e$ as seeds. \check{S}_{t+1} does not take into account cell divisions occurring between t and $t + 1$. To identify cells that may have divided between the two time points, ASTEC searches for an optimal local set of seeds within each cell of \check{S}_{t+1} (local h-minima operator step). \hat{S}_{t+1} is built from the watershed segmentation using these locally-computed seeds, and therefore takes cell divisions into account. Finally, the two estimated segmentations \check{S}_{t+1} and \hat{S}_{t+1} are compared to identify artefacts and produce the final S_{t+1}^* .

4.4 Segmentation projection

For each cell $c \in C_t$ of S_t^* , ASTEC projects S_t^* onto the frame of I_{t+1} to get an estimation of the space occupied by its progeny in I_{t+1} . This process includes 2 steps: i) computation of the non-linear transformation that registers the intensity image I_t onto I_{t+1} ; ii) application of this transformation to S_t^* to map its segmentation onto the frame of I_{t+1} .

The non-linear transformation $\mathcal{T}_{t \leftarrow t+1}$, which maps the I_{t+1} frame into the I_t one (Fig. S10) is computed using the block-matching algorithm described in (50, 51). This transformation is then used to project S_t^* onto the frame of I_{t+1} . However, a direct application of $\mathcal{T}_{t \leftarrow t+1}$ to S_t^* could fail to precisely register cells undergoing fast migrations or shape changes between consecutive time points. To avoid that the projection of a non-dividing cell $c \in C_t$ slightly overlaps two cells in C_{t+1} , each cell $c \in C_t$ of S_t^* is first eroded by applying 10 iterations of a 3D 6-connected structuring element. Eroded cells define regions $\{R_c^e\}_{c \in C_t}$, which collectively form the eroded segmentation S_t^e .

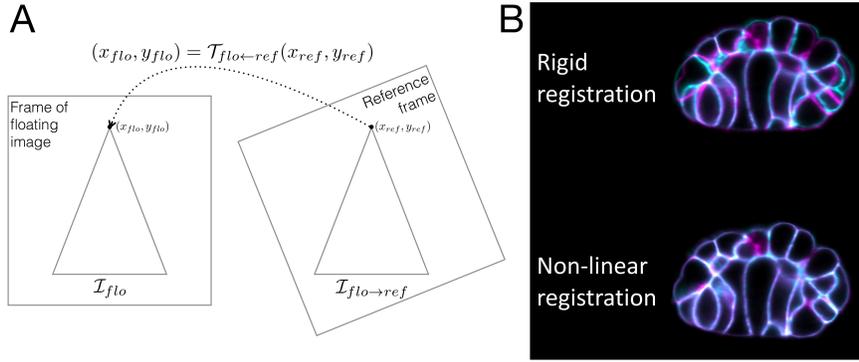


Figure S10: Naming convention and efficiency of the registrations. A) Let \mathcal{I}_{flo} be an image that we want to register into the frame of a corresponding image \mathcal{I}_{ref} (the two images can be consecutive images of the same object or images from different angles for example). Let $\mathcal{T}_{flo \leftarrow ref}$ be a function that maps coordinates from the frame of \mathcal{I}_{ref} to the corresponding coordinates in the frame of \mathcal{I}_{flo} . Let (x_{ref}, y_{ref}) be a set of coordinates in the frame of \mathcal{I}_{ref} corresponding to the coordinates (x_{flo}, y_{flo}) in the frame of \mathcal{I}_{flo} . Then, $(x_{flo}, y_{flo}) = \mathcal{T}_{flo \leftarrow ref}(x_{ref}, y_{ref})$. Then $\mathcal{I}_{flo}(\mathcal{T}_{flo \leftarrow ref}(x_{ref}, y_{ref}))$ is the intensity of the object in the frame of \mathcal{I}_{flo} at the position (x_{ref}, y_{ref}) in the frame of \mathcal{I}_{ref} . Therefore, $\mathcal{I}_{flo} \circ \mathcal{T}_{flo \leftarrow ref}(x_{ref}, y_{ref})$ is the image \mathcal{I}_{flo} registered onto the frame of \mathcal{I}_{ref} . We name this image $\mathcal{I}_{flo \rightarrow ref} = \mathcal{I}_{flo} \circ \mathcal{T}_{flo \leftarrow ref}$. B) XY sections through ASTEC-Pm1 at two consecutive time points ($t = 22$ in cyan and $t = 23$ in magenta) after rigid (top) or non linear (bottom) registration of the 3D images.

S_t^e is then transformed into the frame of $t + 1$: $S_{t \rightarrow t+1}^e = S_t^e \circ \mathcal{T}_{t \leftarrow t+1}$. The regions of R_c^e in $S_{t \rightarrow t+1}^e$ are finally used as seeds for a watershed applied on I_{t+1, σ_2} . The result of this watershed is the propagated segmentation of S_t^* to time point $t + 1$:

$$\check{S}_{t+1} = \mathbf{WS}(S_{t \rightarrow t+1}^e, I_{t+1, \sigma_2}). \quad (4.1)$$

4.5 Cell division detection

\check{S}_{t+1} segmentations present a major source of errors: cell divisions occurring between t and $t + 1$ are not taken into account, leading to undersegmentation (\mathbf{E}_{me}) errors. Second, since the seeds used to build \check{S}_{t+1} are images by the transformation of eroded cells, they may not always include a *bona fide* local intensity minimum. This may bias the watershed and produce cell shape errors, which however should not exceed the size of the erosion. ASTEC therefore next explores whether each cell \check{c} of \check{S}_{t+1} includes one or two sister cells (i.e. includes one or two seeds), and computes optimally determined local seeds to initiate a global watershed.

To identify whether cell \check{c} should be split into two sister cells, ASTEC considers the corresponding region $R_{\check{c}}$ of I_{t+1, σ_1} and computes sets of local h -minima within $R_{\check{c}}$, for all $h \in [h_{min}, h_{max}]$ (in our case $h_{min} = 2$ and $h_{max} = 18$). For each h value, $R_{\check{c}}$ can contain either zero, one, two or more seeds. The aim of this part of the algorithm is to identify for each $R_{\check{c}} \in I_{t+1, \sigma_1}$ an optimal h value, which produces the number of seeds corresponding to the number of cells present in $R_{\check{c}}$.

Since it is not desirable to tune the h range, $[h_{min}, h_{max}]$, for each temporal series, the image I_{t+1} is converted into 8-bit integers beforehand. To do so, the 1 % and 99 % values of the image histogram are respectively mapped to the 0 and 255 values, while values in-between are linearly mapped. Values below the 1 % are set to 0, while values above the 99 % are set to 255.

Because of the high temporal frequency of the acquisitions, cell \check{c} can at most include two sister cells. For low h values the region will tend to be overseeded while for high h values, the number of detected seeds may drop to 1 or even 0 (no local minimum inside the region) (Fig. S11). h values yielding exactly two seeds may either point to a cell division event, or to a high noise level. To discriminate between these two alternatives, we consider respectively the maximal and minimal values of h yielding exactly two seeds and denoted respectively $h_{2+}(\check{c})$ and $h_{2-}(\check{c})$. The length $N_2(\check{c})$ of the interval $[h_{2-}(\check{c}), h_{2+}(\check{c})]$ is:

$$N_2(\check{c}) = h_{2+}(\check{c}) - h_{2-}(\check{c}) + 1. \quad (4.2)$$

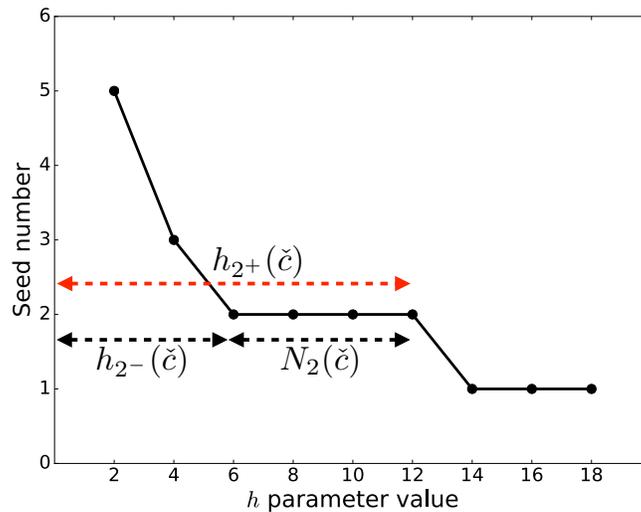


Figure S11: Number of seeds found for a cell \check{c} for different values of the parameter h for seed detection.

We consider that a cell division event is plausible when the two following conditions hold: i) the plateau of h values detecting 2 seeds is large enough (i.e. $N_2(\check{c})$ is large) (Fig. S11) and ii) corresponding h values are high enough to be above the cytoplasm noise ($h_{2+}(\check{c})$). To formalize this, we imposed the following condition for the detection of a cell division in cell \check{c} :

$$s(\check{c}) = h_{2+}(\check{c}) \cdot N_2(\check{c}) \geq \tau, \quad (4.3)$$

where we empirically determined the value of the cell division detection sensitivity parameter $\tau = 25$.

In most cases, multiple values h are valid for the targeted number of seeds and the highest one, $h_{2+}(\check{c})$, is chosen as it corresponds to the largest seeds. Unlike in the MARS algorithm, where a unique h value is used for all cells, the ability of ASTEC to assign a different h to each cell is a crucial element of its performance. Fig. S12 illustrates the distribution of selected h values over the whole embryo sequence. While the majority of cells are segmented with a high h , a small percentage of cell snapshots at each time point were segmented with a small h value.

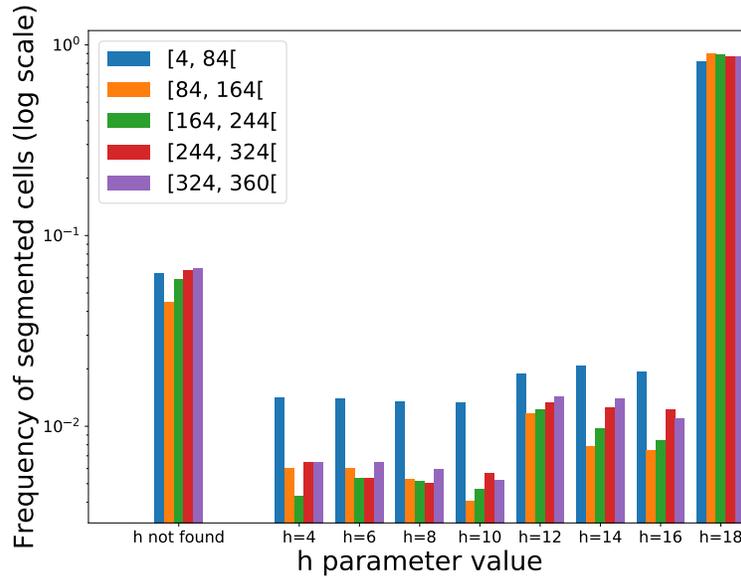


Figure S12: Distribution of h values selected during the ASTEC-Pm1 segmentation and tracking. The developmental sequence was split into 5 temporal segments, and the colors refer to the window of time (in minutes) considered. For some cells, the algorithm was not able to find an h value identifying 1 or 2 seeds within the projected volume of a mother cell. In such case, the cells are sorted as h not found and the projection of the cell is used in the final segmentation.

In less than 3% of cases, the image quality was insufficient to detect either 1 or 2 seeds in \check{c} . This is for example the case for epidermal cells whose apical membrane (facing the exterior of the embryo) was faint or undetectable, precluding seed detection. There were also cases where 3 or more seeds were found for all h values returning a non-null number of seeds. Two heuristics are deployed in order to build coherent segmentation from these rare cases.

Heuristic 1. In the case where 3 seeds (but neither 1 nor 2) are found by the local minima detection at $t + 1$ in the image of a cell at time t , they are first all used to segment \check{c} . Because of the high frequency of acquisition, a cell cannot divide more than once between two consecutive time points. Therefore at least

two - and possibly all - of the three segmented cells at $t+1$ cells must result from the over-segmentation of a single cell. The post-correction process (see below) efficiently resolves the oversegmentation of a cell into two cells, but does not cope with 3-cell oversegmentations. This heuristic therefore aims at fusing two of the three cells to bring the number of oversegmented cells to at most 2. The heuristic chooses two regions that are most likely to result from an oversegmentation. The smallest of the three regions, r_1 , has a volume of at most $1/3$ of the volume of its parental cell. As most cell divisions during ascidian embryogenesis are geometrically equal (Fig S24), r_1 is most likely to result from an oversegmentation event. To decide to which of the other two regions r_1 should be fused, the surfaces of contact between r_1 and the two other regions are compared. We observed empirically that the surface of contact between two oversegmented regions of a cell is supported by noise in the intensity image, and highly irregular. We therefore considered that r_1 shares its largest interface with the region it should be fused to. Fusion of these two regions results in a new segmentation of \check{c} , which only contains 2 cells.

Heuristic 2. In the case where either 0 or more than 3 seeds are found, we either cannot seed the cell (0 seeds) or we cannot decide how to fuse over-segmented parts. In such cases, the eroded cell of \check{c} from $S_{t \rightarrow t+1}^e$ is kept as the seed for the region $R_{\check{c}}$. This is done unless the volume of this eroded cell is smaller than V_{min} in which case \check{c} is considered as a residual of a previous mistake and therefore discarded (in our case V_{min} was empirically assigned the value of 100 voxels or $2.7 \mu\text{m}^3$).

Based on this local analysis, a set of seeds can be determined for each cell region from the locally predicted h values. When put together, this set of seeds forms the image of seeds $Seeds_{t+1}$ necessary for the watershed:

$$Seeds_{t+1} = \mathbf{SR}(\check{S}_{t+1}, \tau), \quad (4.4)$$

where \mathbf{SR} stands for the complete procedure for Seed Refinement described above. The final segmentation of I_{t+1} is then estimated by:

$$\hat{S}_{t+1} = \mathbf{WS}(Seeds_{t+1}, I_{t+1, \sigma_2}). \quad (4.5)$$

Based on these seed detection operations, the progeny set $T(c)$ (for tree) at $t+1$ of each cell region c at time t is constructed for every cell c of the embryo ($T(c)$ is a set of labels from time $t+1$). The cells in $T(c)$ are the cells resulting of the seeds found by ASTEC included in \check{c} , the projection of the cell c . Cell lineages can be tracked at the end of the algorithm using the information stored in T mappings.

4.6 Segmentation consistency checking

\hat{S}_{t+1} contains segmented cells obtained after detection of novel cell divisions. First, all cells with a volume inferior to $MinVol = 10^3$ voxels ($= 27 \mu\text{m}^3$), mostly oversegmented cells, were deleted. Second, we noticed that the identification of a single seed per cell was occasionally insufficient for our watershed algorithm to model correct cell shapes in \hat{S}_{t+1} . Two main error types were identified.

First, a single seed was occasionally insufficient to correctly reconstruct the shape of a single cell. For example, it can happen that the signal that separates two neighbouring projected cells \check{c}_1 and \check{c}_2 is lower than the noise inside one of these cells (\check{c}_1 for example). In that case, if the noise in \check{c}_1 separates that cell into two or more parts, a seed is necessary in each of these parts in order to retrieve the correct shape of that cell. ASTEC might have correctly identified the noise and put only 1 seed in \check{c}_1 instead of the number necessary to recover the whole shape of the cell. We classify this problem as ‘‘High intracellular

noise problem”. In such cases Heuristic 3 below was implemented to add one or more other seeds within the cell, required for an accurate reconstruction of this region.

A second issue occurs in places of the image with a fainter membrane signal intensity, as seen for example in the embryonic epidermal cells (see Fig. S2). In such cases, external cells in \hat{S}_{t+1} may lose material to the benefit of the outside of the embryo, and progressively disappear. This problem, denoted as “shrinking outer cell” problem is handled by Heuristic 4 below.

Both heuristics rely on the fact that, by construction, \check{S}_{t+1} provides a reliable estimate of the shape and volume of the region occupied by the progeny at $t + 1$ of a cell from S_t^* . If a significant difference in volume between a cell in \hat{S}_{t+1} and its progeny in \hat{S}_{t+1} (a single cell or the union of its two daughters in the case of cell division) is detected, ASTEC considers that an error was made during the construction of \hat{S}_{t+1} and the following heuristics are sequentially applied; Heuristic 3 is first applied on the whole embryo followed by Heuristic 4.

Heuristic 3. Correction of “High intracellular noise errors”. If the volume of a segmented cell \check{c} in \check{S}_{t+1} is at least $\tau_1\%$ bigger than its progeny \hat{c} in \hat{S}_{t+1} we assume that \hat{c} was wrongly segmented due to an insufficient number of detected seeds (in our case, we used a conservative value of $\tau_1 = 50$). The correction then consists in using the maximum local value of h increasing the number of seeds in that region. The new number of seeds is therefore strictly greater than the original number of seeds. This new number of seeds can therefore be two or more. If this new number of seeds is 2 then we assume that the algorithm overlooked a cell division and the cell is divided into two daughter cells. We assume that a new number of seeds strictly greater than 2 only occurs in the case of non-dividing cell (because we empirically observed that individual dividing cells usually have a high membrane signal and that they should not exhibit this kind of error). Therefore the resulting regions are fused in order to form one single cell.

Heuristic 4. Correction of “Shrinking outer cell” errors. Cells in \hat{S}_{t+1} that are at least $\tau_2\%$ smaller than their equivalent in \check{S}_{t+1} are tagged for correction (in our case, we used $\tau_2 = 10$). Let \hat{c} be a cell region to correct. We apply an active contour algorithm (Morphosnake algorithm described in (56) and implemented by P. Marquez Neila), using the dilated shape of \hat{c} as the initial contour and the gradient norm transformation of I_{t+1} as the intensity profile. A dilated region is obtained by iterating IT times a 3D 6-connected structuring element on \hat{c} . The morphosnake algorithm is then applied up to stability (at $\pm\delta$ voxels) or after n iterations (in our case, $\delta = 10^3$ voxels and $n = 200$). After this procedure, the voxels belonging to the exterior in \hat{S}_{t+1} that intersect with the shape resulting of the Morphosnake are allocated to the corrected cell $\hat{c}^\#$. A third segmentation $\hat{S}_{t+1}^\#$ is obtained by replacing in \hat{S}_{t+1} all \hat{c} cells by their morphosnake-corrected cells $\hat{c}^\#$. Finally, potential local over-expansion of outer cells due to this previous morphosnake operation are erased by applying an opening operation (using the approximation of a sphere of radius R voxels as a structuring element, $R = 20$) to the whole segmented embryo $\hat{S}_{t+1}^\#$ where only the parts added by this heuristic can be erased. This last operation results in S_{t+1}^* , the final segmentation of time point $t + 1$. As a side note, it can happen that \check{c} from \check{S}_{t+1} corresponds to two cells in \hat{S}_{t+1} . As the intensity of the membrane signal becomes stronger during mitosis, it is unlikely that a cell losing volume to the exterior is dividing. If a dividing cell is found to lose volume to the exterior, it is probably an artefactual division, and the two cells are fused.

The tracking T of the cells from t to $t + 1$ is updated with the newly corrected cells from S_{t+1}^* .

4.7 Post-processing

We found that the segmented series $\{S_t^*\}_{t \in [t_1, t_f]}$ still included segmentation errors. Compared to the two-pass algorithm, the propagative nature of the ASTEC segmentation and tracking algorithm, however, greatly restricts the type of possible errors, and the impact these have on the overall quality of cell tracking (Fig. S13).

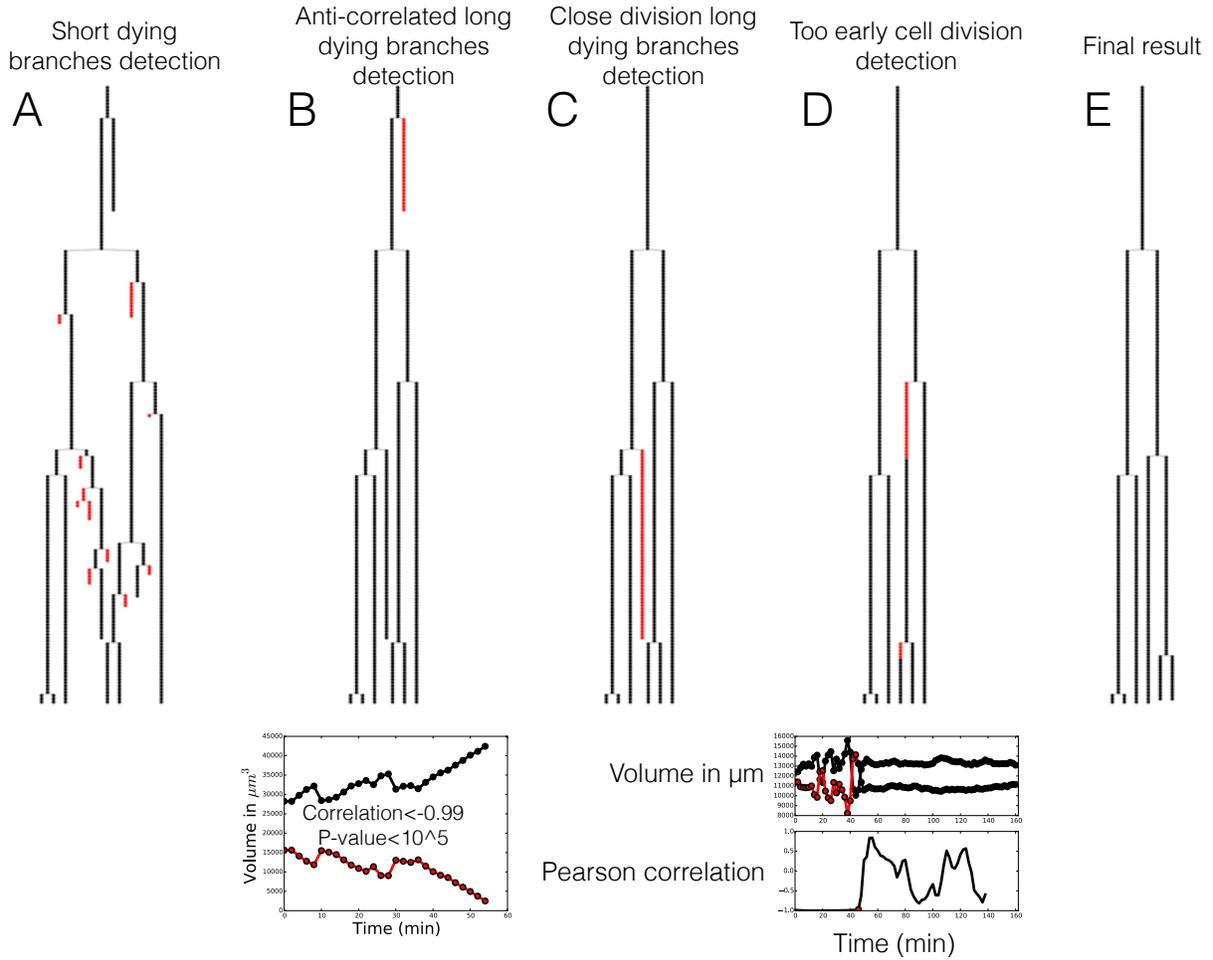


Figure S13: The ASTEC post-correction pipeline. A) Initial output of ASTEC: the lineage contains short interrupted branches (< 25 time points, red), which are first fused with their sister with a longer branch. B) Long interrupted branches (≥ 25 time points) with anti-correlated cell volume to their sister branch are fused with their sister. C) In case of abnormally close cell divisions (< 25 time points), the cells with anti-correlated volumes are fused. D) The precise timing of each cell division is finally checked by looking for anti-correlation between the volumes of sister cells, in which case, the fusion of sister cells postpones the division until the end of the volume anti-correlation period. E) Final corrected cell lineage.

1. “Appearing cell” E_a errors cannot occur, as by design each cell of S_{t+1}^* must have an antecedent in S_t^* . An important corollary of this property is that when an internal coelomic cavity appears in an embryo (e.g. the archenteron), it will not have an antecedent cell and will therefore not be segmented. This situation would be more difficult to manage with a two-pass algorithm.
2. “Missing cell” E_{mi} errors: E_{mi} errors are very rare and only appear in our case when an ectodermal

cell progressively loses material to the profit of the exterior. These errors cannot be transient, as no *de novo* appearing cell can be created, and therefore permanently interrupt a cell lineage, all progeny of the missing cell lacking in the segmented embryo. This type of error cannot be post-corrected, except by artificially prolonging the interrupted lineage tree branch to the last time point (which at best corrects the lineage, the progeny of the missed cell remaining absent from the segmentation).

3. “Merged cell” E_{me} error: these errors are also restricted compared to the two-pass algorithm. Because of the propagative nature of the ASTEC algorithm, it is impossible to fuse two topologically neighbouring cells from different lineages (Fig. S8, E_{me} column). E_{me} errors are therefore restricted to transiently missed cell divisions. This cannot be post-corrected.
4. “Split cell” (oversegmentation) E_s errors can occur, reflecting precocious or artefactual cell division. This type of error, which is the most common type of error, can be efficiently post-corrected by fusing oversegmented cells, as described below.

Post-correction of precocious or artefactual cell divisions takes advantage of the propagative nature of ASTEC and of biological properties of the studied organism:

1. E_s errors are propagated, which implies that they persist over several time points.
2. As ascidian embryos develop without overall growth, ASTEC considers that the volume of an accurately segmented cell c remains roughly constant over successive time points.
3. If a cell c is oversegmented into two cells c_1 and c_2 , the sum of the volumes of c_1 and c_2 should be equal to the volume of c .
4. If a cell c is oversegmented into two cells c_1 and c_2 , then the artefactually-detected membrane separating c_1 and c_2 is due to intracellular noise, which has no reason to remain constant over time. The individual volumes of c_1 and c_2 have therefore also no reason to be stable in time.

Under these assumptions, we expect that the volumes of two cells c_1 , c_2 resulting of the oversegmentation of a cell c evolve in an anti-correlated manner. This expectation, which was empirically validated, is at the heart of the post-correction algorithm.

Since errors only affect a single cell lineage, the post-correction algorithm sequentially and independently analyzes sibling relationships. It first fuses sibling cells with short cell cycle duration, corresponding to short-lived over-segmentations, to their sister (length $< SL$, in our case $SL = 50 \text{ min} = 25 \text{ time points}$, Fig. S13A). Long interrupted branches (length $\geq SL$) could be due either to an artefactual cell division, or to the premature progressive disappearance of an existing cell as time proceeds (E_{mi} error). To discriminate between these two situations, ASTEC compares the volume of the two sisters. If they are anticorrelated (Pearson correlation under -0.9) over the whole length of the branch, the division is considered artefactual and the two sisters are fused (Fig. S13B). The search for anticorrelated volumes can also resolve slightly more complex situations such as an artefactual division shortly preceding a *bona fide* cell division (Fig. S13C). If the volumes of the two potential sister cells are not anti-correlated, we assume that the interrupted lineage results from an E_{mi} error. This anomaly occurs only for 7 cells in our segmentation, and in that case, the lineage branch is prolonged to the final time point with all the properties of the cell (as the volume or the surfaces of contact for example). Note that this correction

does not retrieve any of the cell divisions that may take place downstream of the disappearance of the cell. As a final step, volume anti-correlation is also used to detect and correct precocious cell divisions. For this, sisters with strongly anti-correlated cell volumes (score < 0.8) during a time window of W_s time points or more (in our case, we chose $W_s = 10$) starting with their birth are fused over the period of anticorrelation (Fig. S13D). The corrections applied to ASTEC-Pm1 during post-processing are listed Table S2: the post-correction step fused over 7700 cells and efficiently corrected all but one of the E_s errors.

The efficiency of E_s error post-corrections indirectly provides a way to reduce the number of E_{me} errors, which cannot be post-corrected. By selecting a low value for the cell division detection sensitivity parameter τ of equation 4.3, ASTEC can be biased towards the detection of artefactual cell divisions (E_s), which can be efficiently corrected, over under-segmentation (E_{me}), which cannot.

4.8 Additional module: region-based membrane enhancement

Two difficulties may arise when processing image sequences for the segmentation propagation: cell membrane contrast may become inhomogeneous in the fusion stack across time (see Fig. S14, left panel) and some membrane signal may become weak. This may compromise the quality of the segmentation obtained at $t + 1$ by the segmentation propagation procedure. To alleviate these defaults, the fusion image I_{t+1} is then pre-processed prior to the segmentation propagation by taking advantage of the segmentation S_t^* obtained at the previous time point t .

The above mentioned difficulties being spatially inhomogeneous in the fusion image, pre-processing operations at time $t + 1$ are done on a regional basis. More precisely, this segmentation is projected onto the frame of I_{t+1} by $S_t^* \circ \mathcal{T}_{t \leftarrow t+1}$, cells $\circ S_t^* \circ \mathcal{T}_{t \leftarrow t+1}$ are used to define sub-images in I_{t+1} where pre-processing operations are conducted independently and then fused afterwards to get the whole pre-processed image.

Two pre-processing operations were designed. To circumvent the first difficulty, a dedicated membrane enhancement was designed. It is a region-based version of the tensor voting enhancement described in (57) that is applied to the fusion image I_{t+1} . The second difficulty is addressed by a region-based intensity 1-byte conversion of the fusion I_{t+1} . Both resulting images are finally combined by the maximum operation (see Fig. S14, right panel), and this last image replaced the fusion image I_{t+1} for the segmentation propagation procedure.

This module was used to segment and track all datasets described in this article, except ASTEC-Pm1.

4.8.1 Tensor voting based membrane enhancement

First, a valued image E_{t+1} of the membrane centerplanes is computed. At each voxel \mathbf{v} , the Hessian matrix is computed (by convolution with a Gaussian of $\sigma_{memb} = 0.9 \mu\text{m}$). Its eigenvector $\mathbf{u}_3(\mathbf{v})$ associated with the largest eigenvalue $|\lambda_3(\mathbf{v})|$ indicates the orthogonal direction of a potential membrane. It allows to compute a response image R_{t+1} by

$$R_{t+1}(\mathbf{v}) = \frac{1}{2} (\nabla I_{t+1}(\mathbf{v} - \sigma_{memb} \mathbf{u}_3(\mathbf{v})) \cdot \mathbf{u}_3(\mathbf{v}) - \nabla I_{t+1}(\mathbf{v} + \sigma_{memb} \mathbf{u}_3(\mathbf{v})) \cdot \mathbf{u}_3(\mathbf{v}))$$

Directional extrema of $R_{t+1}(\mathbf{v})$ in the direction of $\mathbf{u}_3(\mathbf{v})$ yield the membrane centerplane image E_{t+1} .

Second, to deal with inter-cell contrast variability, this centerplane image is thresholded on a cell-based basis. The bounding box of each cell c is calculated, then dilated by $3.6 \mu\text{m}$, and calculations take place in the extracted sub-images E_{t+1}^c . So-called directional histograms of E_{t+1}^c are computed, it allows to handle the anisotropic point spread function of the microscope. Recall that $\mathbf{u}_3(\mathbf{v}) = (u_x(\mathbf{v}), u_y(\mathbf{v}), u_z(\mathbf{v}))$ is an unit vector, thus verifying $u_x^2(\mathbf{v}) + u_y^2(\mathbf{v}) + u_z^2(\mathbf{v}) = 1$. The directional histogram H_d^c for $d \in \{x, y, z\}$ is computed by:

$$H_d^c(e) = \sum_{\mathbf{v}|E_{t+1}^c(\mathbf{v})=e} u_d^2(\mathbf{v})$$

The shape of the directional histograms $H_d^c(e)$ suggest they can be approximated by the analytic function $\frac{a}{e^b} + c \frac{e}{\sigma^2} \exp^{-e^2/(2\sigma^2)}$, the first term representing the response to noise while the second one (a Rayleigh distribution) is considered as the contribution of the membrane signal. Parameters are computed by minimization with the Levenberg-Marquardt algorithm. Directional thresholds t_d^c are set to be the first 1 % of the computed Rayleigh distribution, which allow to compute a direction-adaptive threshold $t(\mathbf{u}_3) = u_x^2 t_x^c + u_y^2 t_y^c + u_z^2 t_z^c$. The binarized E_{t+1}^c are masked by the cell shape dilated by $3.6 \mu\text{m}$, and all masked and binarized E_{t+1}^c are gathered into a single binary image B_{t+1}^c .

Last, a reconstructed membrane image is computed from the B_{t+1}^c thanks to the tensor voting framework of (58). Each non-null voxel \mathbf{v} of B_{t+1}^c is supposed to belong to a membrane orthogonal to $\mathbf{u}_3(\mathbf{v})$ and can be represented as a stick tensor. Perceptual grouping is achieved by collecting the voting fields (defined by $\sigma_{TV} = 3.6 \mu\text{m}$) of all these stick tensors. A surfaceness image is computed from the resulting tensor field (57). This is followed by a Gaussian smoothing with $\sigma = 0.9 \mu\text{m}$). Since the tensor field is normalized, the resulting image is in the range $[0, 1]$. It is converted into 1-byte after multiplying by 255 (see second image from the left in Fig. S14).

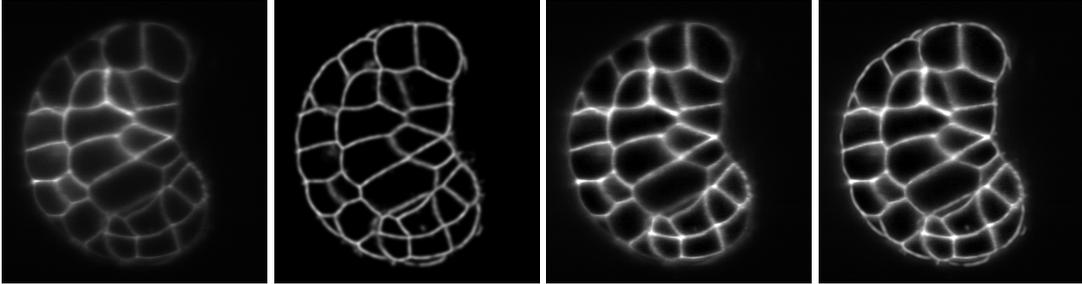


Figure S14: From left to right: an example *XY*-section of a reconstructed fused image; cell-based membrane enhancement with tensor voting; cell-based intensity 1-byte conversion; fusion of the two preceding images.

4.8.2 Intensity 1-byte conversion

This cell-based intensity 1-byte conversion is inspired by the 1-byte conversion described in section 4.5 (page 19). The rationale of this 1-byte conversion is to stretch the useful intensity range into $[0, 255]$, implicitly assuming that the maximal intensities will lie in the membrane signal, while the minimal intensity will be either in the cell interior or in the background. Its cell-based version consists in computing, for each cell c a maximum value $max(c)$ from only the cell border (defined as the union of the inner and outer 6-boundaries) and a minimum value $min(c)$ from only the cell interior (the cell minus the inner 6-boundary). Extremum values are computed from the histograms of the respective areas, the maximum and minimum values corresponding respectively to the 99 % and 1 % values of the histograms, similarly to the procedure described in section 4.5. A maximum value image $I_{max}(\mathbf{v})$ is built by filling each cell (including the background considered as a cell) by its corresponding maximum value $I_{max}(\mathbf{v}) = max(c), \mathbf{v} \in c$.

Obviously, intensity jumps will occur at cell boundaries. Therefore, the maximum value image $I_{max}(\mathbf{v})$ is smoothed with a Gaussian kernel ($\sigma = 5$ voxel). A minimum value image $I_{min}(\mathbf{v})$ is built similarly. Last, the intensity $I_{t+1}(\mathbf{v})$ is converted into $[0, 255]$ by $\frac{255}{I_{max}(\mathbf{v}) - I_{min}(\mathbf{v})} (I_{t+1}(\mathbf{v}) - I_{min}(\mathbf{v}))$ (see third image from the left in Fig. S14).

4.9 ASTEC parametrization

In this section, we recapitulate the various parameters that can be adapted to individual datasets and provide the values used in the case of the ASTEC-Pm1-10 *Phallusia mammillata* embryo segmentations.

1. *Initialization*: for the segmentation of the first time point of our dataset, we used MARS with the following Gaussian smoothing parameters: $\sigma_1 = 0.6 \mu\text{m}$ and $h = 4$ for seed detection, $\sigma_2 = 0.15 \mu\text{m}$ for the seeded watershed.
2. *Iteration-projection*: cell erosions were done using a 3D 6-connected structuring element of 1 pixel, iterated 10 times. The same σ_2 parameter as in the initiation step was used during watershed.
3. *Iteration-cell division detection*: the range of h values used for seed detection for this dataset was $H = \{2n \mid n \in [1 \dots 9]\}$. The τ value used to identify cell divisions was $\tau = 25$. The minimal size of a seed to be used in the ensuing watershed is $V_{min} = 2,7 \mu\text{m}^3$.
4. *Consistency checking*: all cells smaller than $MinVol = 1000$ voxels were deleted. The dilations were done using a 3D 6-connected structuring element of 1 pixel, iterated $IT = 10$ times. The morphosnake algorithm was run until reaching equilibrium (at $\pm\delta = 10^3$ voxels or $27 \mu\text{m}^3$) or $n = 200$ iterations. The opening operation was done with the approximation of a sphere with a radius of 20 voxels ($6 \mu\text{m}$).
5. *Post-processing*: the cell lineage branches were considered long when larger or equal to the SL threshold, set at 25 time points. The window size W_s for the sister volume anti-correlation set at 10 time points.
6. *Region-based membrane enhancement additional module*: two membrane enhancement parameters $\sigma_{memb} = 0.9$ for the membrane detection and $\sigma_{TV} = 3.6$ for the tensor voting operation. Increase one or both can be useful to recover missed cells but can dramatically increase the number of over-segmentation.

The segmentation propagation is first launched without this additional module. However, If an outer membrane is interrupted or too faint, a cell may invade the background during the watershed step, and this causes the background to "disappear" after a few more iterations, which, in turn, may cause the segmentation propagation to abort. By convention, cells are numbered with labels greater or equal to 2, while the '1' label designed the background. If the 1-cell (i.e. the background) is observed to disappear, the time point where the background invasion began can be identified by looking backwards to the previous segmentations. It is then advised to restart the propagation one time point before the cell invasion was first observed, either by activating the region-based membrane enhancement module, or by tuning its parameters if already active.

Access to examples of parameters used to process a given movie are indicated in section 14.

4.10 Time of execution

The software was run on a workstation with Debian GNU/Linux 9.4 operating system, with the following features: 64Gb RAM, 40-core Intel Xeon processor, 2.2GHz. The following indications are given for the longest movie ASTEC-Pm1 (192 time points).

- **Fusion:** 24 hours. Total raw image size (4 stacks per time point) before fusion was 1Tb. Cropping, fusing and resampling to a voxel size of $0.3 \times 0.3 \times 0.3 \mu\text{m}^3$ reduced the total 4D image size to 83Gb (on average 432Mb/time point).
- **Initialisation / Mars segmentation:** a few minutes, followed by manual curation of over-segmented cells and naming of cells.
- **ASTEC segmentation:** about 5 days using the additional membrane enhancement modules. The time needed for the segmentation of a single time point can be quite variable and depends on the characteristics of the image.
- **Post-correction:** half an hour.

The segmentation can be accelerated by reducing the size of the fused images by resampling to a voxel size of $0.6 \times 0.6 \times 0.6 \mu\text{m}^3$. This reduces the propagation of the segmentation to 5-10 minutes per time point.

5 Cell naming and extraction of geometrical data

5.1 Naming of cells and fate allocation

Early ascidian development is known to be stereotyped. Each cell can thus be assigned a unique name, which identifies it without ambiguity across different embryos, according to the Conklin nomenclature (59). Briefly, each cell was named $Lp.q$ (e.g. $A4.2$), where L indicates the founding lineage at the 8 cell stage, from which the cell originates: A is anterior vegetal, a anterior animal, B posterior vegetal and b posterior animal, p is the rank of the mitotic cell cycle (e.g. $p = 3$ means that the cell is in its 3rd mitotic cell cycle since fertilization) and q is an integer reflecting the relative position of the cell with respect to the vegetal pole of the embryo, compared to its sibling. During division, cell $Lp.q$ gives rise to two daughters $L(p+1).(2q)$ and $L(p+1).(2q-1)$, where the apical surface of $L(p+1).(2q-1)$ is geodesically closer to the vegetal pole of the embryo. To distinguish left and right cells (as seen from the vegetal pole), left cells were given a * suffix (e.g. $A7.4^*$), while matching right cells were identified by a _ suffix (e.g. $A7.4_-$).

The strategy employed to accurately name all cells in each embryo is a combination of computational implementations of standard naming strategies for ascidian embryos (59, 60), inter-embryonic cell name matching between embryos and manual curations. To begin with, we chose a representative embryo (ASTEC-Pm8) with high bilateral symmetry and developmental quality. Segmented cells produced by ASTEC for this embryo were automatically named based on a variant of the Conklin nomenclature (CN) up to the early gastrula stage. This variant uses an Euclidian distance between cell barycenters to compute the distance of a cell to the vegetal pole, defined as the middle point of $Bx.I^*$ and $Ax.I_-$, which is algorithmically easier and less ambiguous to define than a geodesic distance on the embryo. Because of slight differences in the geometry of embryos of different ascidian species, the strict application to all

species of the Conklin nomenclature, developed for *Styela partita* can sometimes lead to homologous cells in the embryos of different species receiving different names. The automatic naming was therefore manually checked and corrected to fit literature descriptions and to ensure consistency between the names of bilaterally-equivalent cells on either side of each embryo and between embryos up to the early gastrula stage.

Beyond the onset of gastrulation, Conklin's nomenclature becomes too imprecise to be automatically implemented, as the vegetal pole invaginates first, making it an unreliable reference point. To name cells beyond the early gastrula stage, we thus conceptually extended the local logic used by Conklin's to name cells up to the early gastrula stage.

As an example, this strategy was previously adopted by Meinertzhagen and colleagues (60) to name neural plate cells beyond gastrulation. They noticed that up to the gastrula stage, neural cells divide anterior-posteriorly, and after each division the more anterior sister cell received the higher index. They therefore continued using this rule beyond gastrulation, which led to naming the most anterior neural cells with a higher index. We followed the same rule.

This strategy of extension of local rules was extended to other tissues. For instance, following a pre-gastrula cell division in the epidermis, the sister closest to the animal pole (the intersection between the separation line between a-line and b-line epidermal cells, and the midline) generally receives the higher index according to Conklin. Whenever possible we followed this local rule beyond gastrulation, making bilateral ad hoc decisions when the orientation of divisions was such that the distance of both sisters to the animal pole was too similar to be used as a nomenclature cue.

Next, we used a maximal-similarity strategy to project the curated naming of ASTEC-Pm8 cells onto all other embryos. Cells of a generic embryo ASTEC-Pm_i are first manually named at the 64-cell stage. Exploiting the lineage information, the strategy follows in time each of these initial 64 cells up to their division. The neighbourhoods of each of the two daughters $d_i^{(1)}, d_i^{(2)}$ (see below the description of the extraction of these data) are then analyzed throughout their cell cycle and compared to the corresponding neighbourhoods of the same two daughter cells $d_8^{(1)}, d_8^{(2)}$ in ASTEC-Pm8. Such a comparison \mathcal{S} counts the number of conserved neighbours between one cell in ASTEC_i and the same cell in ASTEC-Pm8. When one neighbour is not yet named, the comparison is performed based on the name of its mother cell. Specifically, the strategy calculates $\mathcal{S}(d_i^{(1)}, d_8^{(1)}) + \mathcal{S}(d_i^{(2)}, d_8^{(2)})$ and $\mathcal{S}(d_i^{(1)}, d_8^{(2)}) + \mathcal{S}(d_i^{(2)}, d_8^{(1)})$ and names $d_i^{(1)}$ as $d_8^{(1)}$ (resp., as $d_8^{(2)}$) if $\mathcal{S}(d_i^{(1)}, d_8^{(1)}) + \mathcal{S}(d_i^{(2)}, d_8^{(2)}) > \mathcal{S}(d_i^{(1)}, d_8^{(2)}) + \mathcal{S}(d_i^{(2)}, d_8^{(1)})$ (resp., if $\mathcal{S}(d_i^{(1)}, d_8^{(1)}) + \mathcal{S}(d_i^{(2)}, d_8^{(2)}) < \mathcal{S}(d_i^{(1)}, d_8^{(2)}) + \mathcal{S}(d_i^{(2)}, d_8^{(1)})$). Finally, the strategy is iterated for each subsequent division until the end of the movie. The naming thus produced for embryo ASTEC_i is finally manually checked and, when needed, corrected to respect extended versions of Conklin's and Meinertzhagen's rules.

Based on cell names, the larval tissue fate of each cell, as described in the ANISEED database (61) was manually assigned up to the early gastrula stage, when the fate of most cells is restricted to a single larval tissue type, and subsequently propagated through the ASTEC cell lineages (Fig. S15). This also allowed to track the position in space of cells contributing to a given larval fate (Fig. S16).

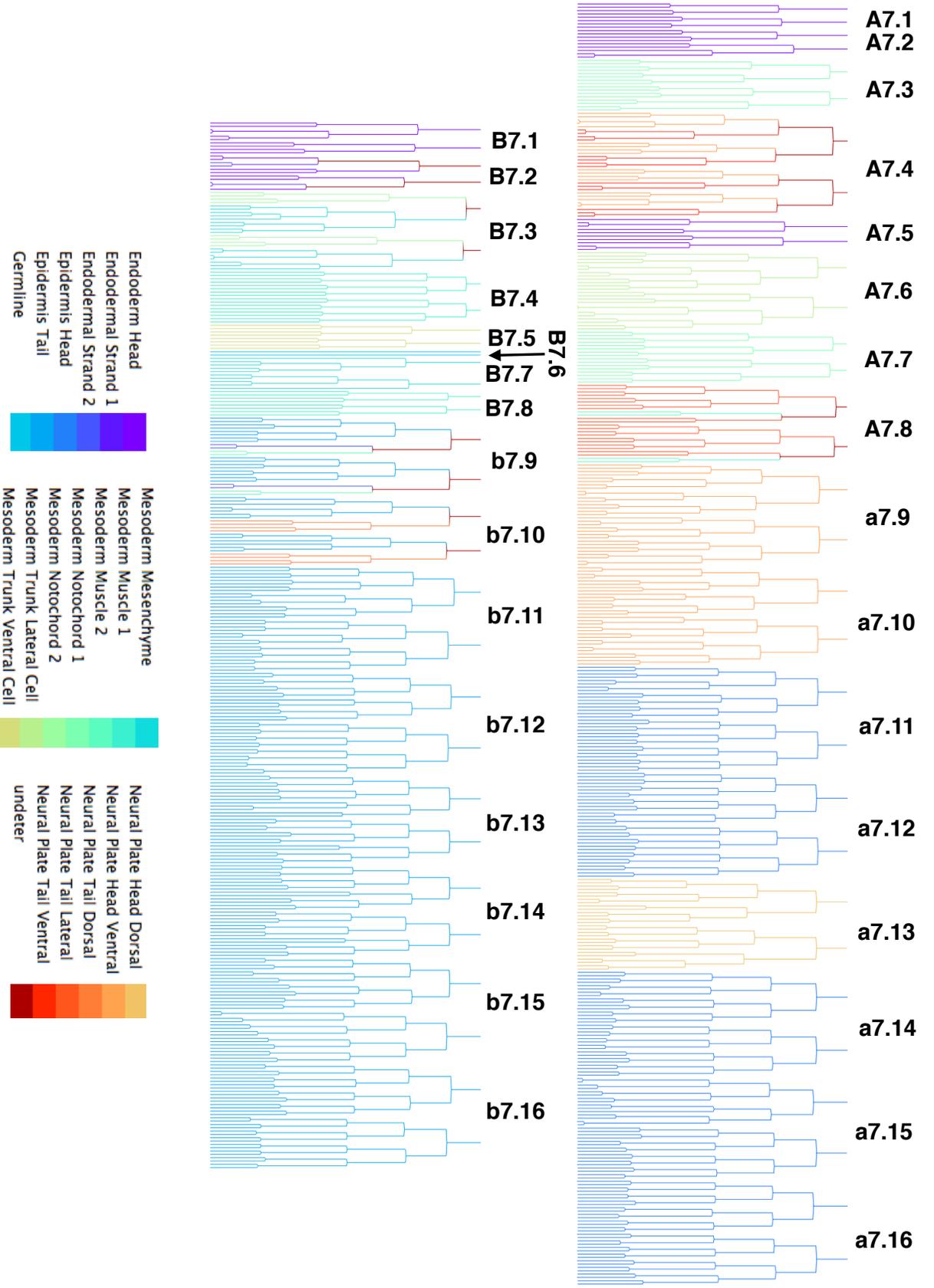


Figure S15: ASTEC-Pm1 cell lineage tree with fate map. For each cell identity, the two corresponding bilaterally symmetrical lineage trees are shown next to one another. Colors indicate the fate of the corresponding sublineage.

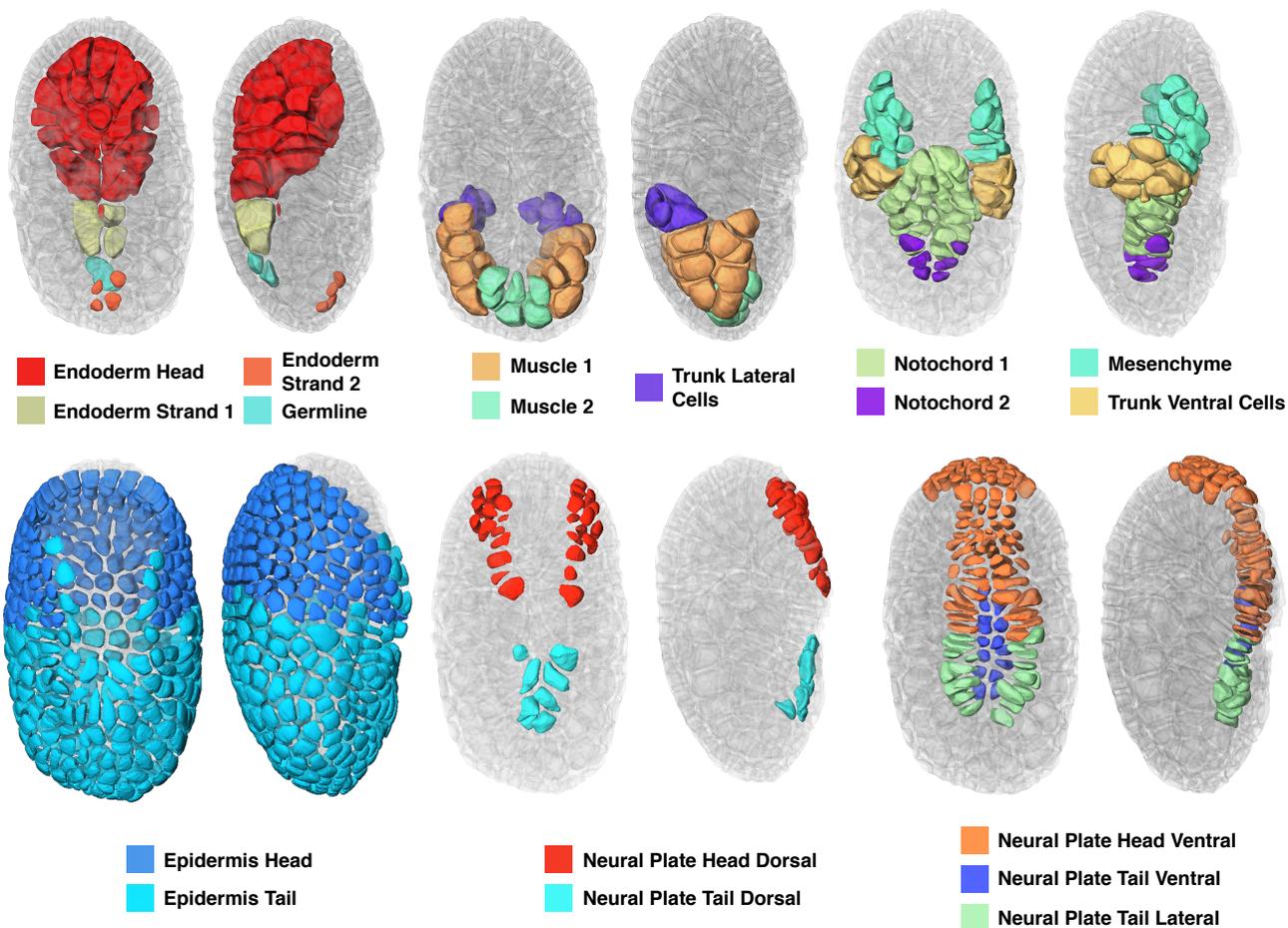


Figure S16: Position of individual tissue precursors at the early tailbud stage in ASTEC-Pm1. In these 3D projections, cells are colored according to their larval tissue fate with the same the color code as in Fig. S15. Segmented cells are slightly eroded to better identify individual cells. Cells not belonging to the tissues of interest are in transparent grey. For each fate, a dorsal (left) and a lateral (right) view of the embryo are shown.

5.2 Organization and extraction of geometrical data

The geometrical information for all segmented cells is stored in a pickle for python (object serialization, .pkl file), and can be extracted using the Jupyter Notebook ASTEC-JupyterNotebook.ipynb provided (see section 14 of supplemental information to access this Notebook and a description of the structure of the .pkl file).

The main features extracted in this piece of work for every cell snapshot at each time point are the total cell volume and surface, and the area of contact between each cell snapshot and its physical neighbours (see ASTEC-JupyterNotebook.ipynb for a full description of the data structure and how to use it).

The volume is defined as the number of voxels belonging to a given segmented cell snapshot. The surface of contact of a cell snapshot c to a neighbour c' is computed with the method described in (62). The total surface of a cell snapshot is calculated as the sum of the surfaces of contact with its neighbours plus its surface of contact with the exterior of the embryo.

Prior to volume and surface computations, the segmentation S^* is smoothed. Indeed, the fusion image may not exhibit signal maximum at membrane centerplane: the watershed-based segmentation may then exhibit noisy borders which, in turn, may lead to an overestimation of the surface computed (see Fig. S17, left and middle panels). Smoothing the segmentation image reduces this issue. The image S^* is first transformed in a set of binary image $B_c(\mathbf{v})$, one par cell c :

$$B_c(\mathbf{v}) = \begin{cases} 1 & \text{if } \mathbf{v} \in c \\ 0 & \text{else} \end{cases}$$

The B_c images are smoothed with a Gaussian kernel ($\sigma = 3$ voxel) to yield the SB_c images. The smoothed cell segmentation image SS^* (see Fig. S17, right panel) is then defined by $SS^*(\mathbf{v}) = \arg \max_c SB_c(\mathbf{v})$.

The compactness \mathcal{C} of a cell is computed as follow:

$$\mathcal{C} = \frac{\sqrt[3]{\mathcal{V}}}{\sqrt[2]{\mathcal{S}}} \quad (5.1)$$

where \mathcal{V} is the volume of the cell and \mathcal{S} is its surface.



Figure S17: From left to right: one XY-section of a reconstructed fused image; its segmentation into cells; the smoothed cell segmentation.

6 Manual assessment and curation of ASTEC-segmented embryos after post-correction

Following post-correction, residual over-segmentation errors were manually corrected, and residual segmentation and tracking errors detected (Table S3).

6.1 Assessment of the quality of the automated segmentation and tracking of ASTEC-Pm1

To assess the automatic segmentation and lineages developed and tested for embryo ASTEC-Pm1, a manually-curated set of 2D and 3D segmented embryos was constructed using the AMIRA v5.4.1 software. For this, the segmented cells of ASTEC-Pm1 were overlaid with the fused membrane fluorescence images and the brush tool was used to add or remove voxels where the ASTEC segmentation did not match the cell contours defined by the fused fluorescence data.

This procedure was carried out on all the cells of the 3D image at time $t = 152$ (218 cells) of ASTEC-Pm1. Each cell was individually corrected, plane by plane, along 3 different views (comparison to ground truth in Fig S6). Additionally, 2D manual curation of a set of 150 2D sections through the embryo covering the whole duration of the movie was performed. Cell detection was assumed to be accurate and all 2D cell snapshots present on the individual sections were individually, manually corrected for their shape using Amira 5.4.1 software (total cell number: 8825).

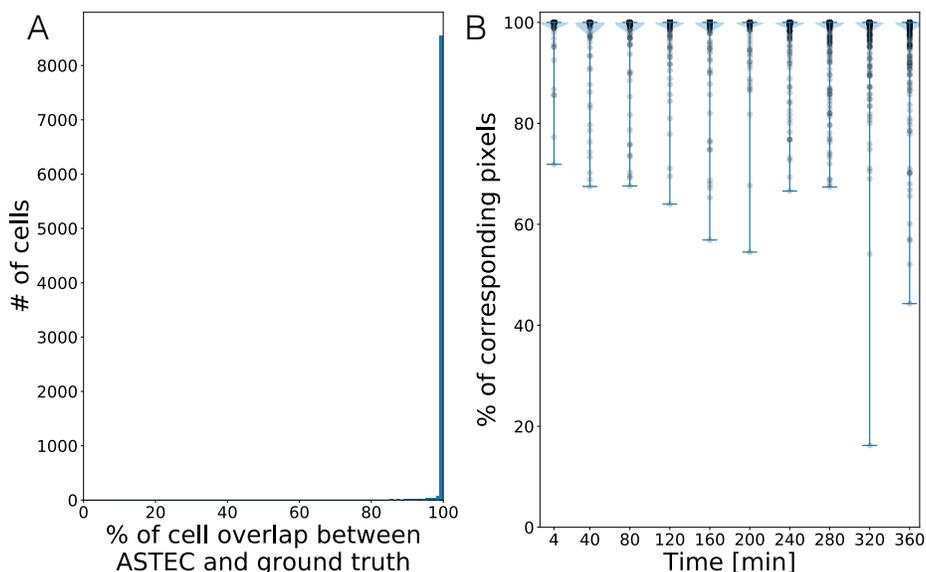


Figure S18: Quantification of the quality of the shape of ASTEC-segmented cells in ASTEC-Pm1. A) Distribution of the dice score \mathcal{D} between corresponding cells c_1 and c_2 from the ground truth and the automatic segmentation ($\mathcal{D}(c_1, c_2) = \frac{2\mathcal{V}(c_1 \cap c_2)}{\mathcal{V}(c_1) + \mathcal{V}(c_2)}$). 100 bins are shown. B) Violin plot of the distribution of the dice scores between corresponding cells. For both plots, the dice scores are computed on a set of 150 2D sections through the embryo covering the whole duration of the movie.

Fig. S18 illustrates that less than 2% of ASTEC-Pm1 segmented cells had Dice score values $< 95\%$ when compared to the 2D curated cells with the matching curated segmented cell. 91% of the cells were found without any detectable error. The 2% of cells with shape issues (Dice score $< 95\%$) were evenly distributed across developmental stages.

6.2 Comparison with RACE, a state-of-the-art segmentation algorithm for cell membrane segmentation and tracking

To assess the performance quality of ASTEC we compared the segmentation accuracy of ASTEC with RACE (Real-time Accurate Cell-shape Extractor). RACE is a state of the art image analysis framework adapted for fast extraction and tracking of cell shape information from large data sets of membrane fluorescent data alone or combined information of membrane and nuclei fluorescent data (16). RACE in a first step segments three-dimensional (3D) image data section by section. In a second step, RACE combines the segmented cells across sections based on similarity and guided by seeds computed for each section. RACE has been shown to produce high quality segmentation in data from *Drosophila*, zebrafish and mouse embryos and is reported to outperform ACME, EDGE4D and MARS (16).

The time lapse movie ASTEC-Pm1 was segmented with both ASTEC and RACE. The parameters for the ASTEC segmentation are described in section 4.9. In order to build the best segmentation possible with RACE and allow a fair comparison between the two methods, we first switched the median filter to a gaussian filter. We then explored more than 30.000 parameter combinations and selected the combination that generated the best overall segmentation across the whole time series. The segmentations were scored according to measurements of agreement to our ground truth. Our measurements of agreement were performed on a subset of time-points throughout the time-series: time-points 1, 90 and 180. We choose the parameter-set that minimized the following scores: segmented areas, boundary detection and fusion quality. A single time point RACE segmentation optimization with a smaller parameter space exploration was also performed, in which the fused image was downsized by a factor of 4 to be compatible with RACE processing. We qualitatively compared the best possible time series output of RACE and the single time point RACE segmentation optimization against the output of ASTEC. The quality of the RACE segmentation output was markedly inferior to ASTEC and to the quality of the segmentation outputs for other species analyzed by (16) (Fig. S19).

The most common problems were caused by under-segmentations during the first step of the process: the independent segmentations of each slice of the image. The second step, whose goal was to aggregate the segmentations of the different slices, was therefore compromised because of these early mistakes. The presence of under-segmentations in a section leads to loss of cell identity maintenance throughout image stacks (see Fig. S19. A, xz view). The central cause for the problems presented by RACE in the analysis of ascidian membrane datasets is most likely, mainly, due to the particular signal features presented by our fluorescent membrane reporter. The membrane signal in our datasets is inhomogeneous in space and time. In space, outer membranes were systematically less labelled than inner membranes leading to under-segmentations (Fig. S19.B). In time, cell diameters decreased by around 3 fold and cell numbers increased by about 8 fold, thereby changing the sample's optical properties (light penetration, scattering) and making the signal inhomogeneous over time.

It was therefore not possible to find a set of parameters that performs well throughout the time-series. RACE is not adapted to process data exhibiting the feature particularities presented by the PH-reporter and the organism used in this manuscript. We additionally compared the segmentation output of ASTEC to an optimized single time point RACE segmentation. The segmentation quality of a single time point optimization was markedly superior to a time series parameter optimization. However, this RACE segmentation was still qualitatively inferior to the ASTEC segmentation output. The ASTEC pipeline thus outperforms both MARS (Fig. S6) and RACE (Fig. S19) in membrane-based cell segmentation of long time-series of developing ascidian embryos.

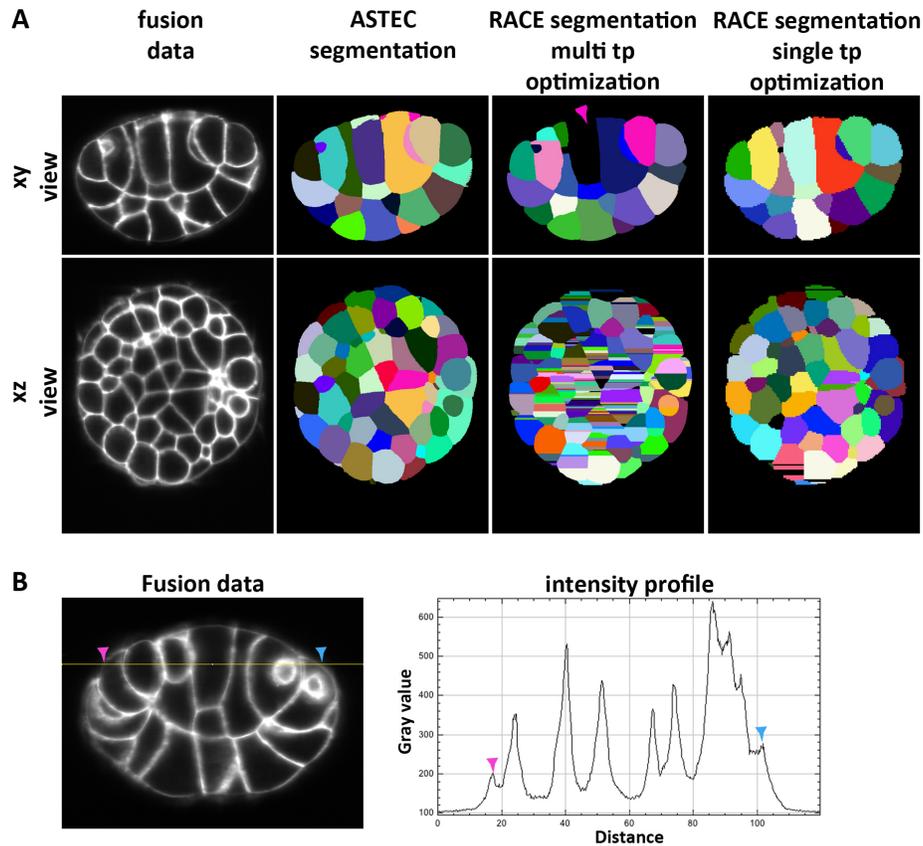


Figure S19: ASTEC generates higher quality ascidian segmentation outputs than RACE. A) Single section of isotropic fused fluorescence membrane input data and of segmentation outputs produced by ASTEC, RACE optimized for multiple time points (multi tp optimization) and RACE optimized for a single time point (single tp optimization). The xy view represents the orientation of RACE section by section segmentation processing. Under-segmentation errors in RACE are highlighted by arrow heads in panel A. The xz view represents an in depth cut (from a top/xy view) in the direction of the RACE section to section association. This view highlights errors in section to section association by RACE. B) Fluorescence signal profile along a line in a single section. The points highlighted by arrowheads showcase the lower intensity of the membrane signal at the edge of the embryo.

7 Analysis of the stereotypy of *Phallusia mammillata* development (Supplement to Figure 2)

Ascidians are known for the stereotypy of their early development, first described by Conklin at the beginning of the 20th century. To go beyond previous qualitative analyzes and quantitatively assess the stereotypy of *Phallusia* development beyond the cleavage stages, we define below several metrics used to compare ASTEC-Pm1-10. We distinguished temporal and spatial stereotypy. Temporal stereotypy was assessed by comparing the evolution in time of cell numbers and cell cycle durations and the lineage tree structure across datasets. Spatial stereotypy was assessed by comparing the geometrical and topological properties of single cells.

7.1 Temporal stereotypy

Precision of cell division detection. To analyze the precision of the timing of cell divisions, we reasoned that during mitosis, mother cells round up to produce two initially roundish daughter cells. Consistently, the compactness \mathcal{C} of cells first increased then decreased during mitosis (Fig. S20, S21). To achieve a high-resolution view of the *circum*-mitosis period, we aligned all cell divisions from a given embryonic cell cycle and plotted the evolution of compactness at each time point in a window of 20 minutes centered on the detected end of cytokinesis (Fig. S20). As expected, at the end of cell cycle 7, compactness first increased up to within 4 minutes of the completion of cytokinesis, then decreased during cytokinesis. The two daughter cells were initially very compact, before progressively adopting a less compact interphase shape. A similar pattern was observed for later cell cycles, except that the peak of compactness of the mother cell was slightly shifted towards the end of detected cytokinesis, and the decrease of daughter compactness was shifted away from detected cytokinesis. We conclude that during these later time points, ASTEC may have prematurely detected cytokinesis, but that this error was on average not larger than 2 minutes, or one time point.

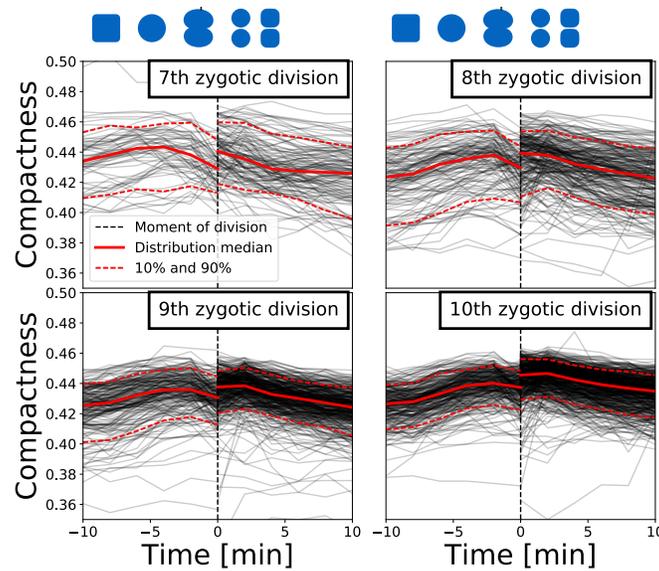


Figure S20: Evolution of cell compactness around mitosis in *ASTEC-Pm1*. Cell divisions at the end of the indicated cell cycle were temporally aligned onto the vertical black hashed line. Overlying schema exemplify expected shape changes. The red line represents the median of the distribution, the red hashed lines the bottom and top 25% of the distribution.

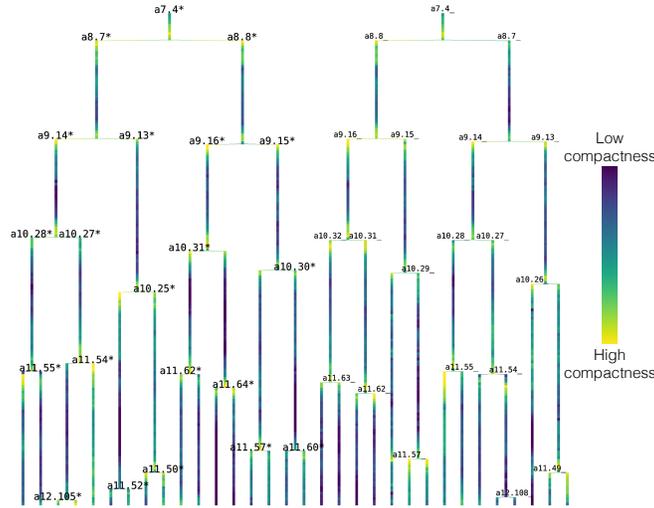


Figure S21: Comparison of ASTEC cell lineage trees from bilateral A7.4 blastomeres from ASTEC-Pm1. Node color represents the compactness of the cell at this time point, from dark blue, not compact, to yellow, compact. Note that the cells are more compact just before and after their division. For the sake of clarity, not all cell names are displayed. Tree visualization using Tulip 5 (63)

Temporal registration of ASTEC-Pm1-10 datasets and evolution of embryonic cell numbers. To analyze how stereotyped the temporal development of *Phallusia mammillata* embryos is, we compared the increase in cell number over time in the ten segmented embryos ASTEC-Pm1 to ASTEC-Pm10. As these embryos were imaged at slightly different temperatures, we first temporally registered their development, using ASTEC-Pm8 as reference. For all embryos, the analysis of cell numbers suggested a linear time scaling between embryos. We identified the optimal scaling factors minimizing the difference in cell numbers between reference and each other embryo. By doing so and using the information that the time interval between two subsequent time points of ASTEC-Pm8 is $\Delta t_{A8} = 120$ s, we find $\Delta t_{A1} = 88$ s, $\Delta t_{A2} = 122$ s, $\Delta t_{A3} = 107$ s, $\Delta t_{A4} = 106$ s, $\Delta t_{A5} = 100$ s, $\Delta t_{A6} = 114$ s, $\Delta t_{A7} = 100$ s, $\Delta t_{A9} = 114$ s and $\Delta t_{A10} = 116$ s. Following this linear temporal rescaling, the evolution of cell numbers in these datasets are in remarkable agreement (Fig. 2B, S22C), suggesting that *Phallusia* development is essentially proceeding with fixed cell numbers. These 10 temporally-aligned embryos were used for all subsequent natural variability analyzes.

Cell cycle duration comparison. We define the cell cycle duration (CCD) comparison metric \mathcal{M}_L as the relative CCD variation among matching cells. Take A and B as two matching cells, and be L_A and L_B the time elapsed between the division of their mother and their own division, measured in hours after the temporal registration procedure detailed above. We define

$$\mathcal{M}_L(A, B) = \frac{|L_A - L_B|}{L_A + L_B}. \quad (7.1)$$

Also in this case the measure lies in $[0, 1)$, with 0 corresponding to perfectly equal cell cycle durations. Fig. S22 presents a quantification of the degree of stereotypy in the cell cycle durations of intra- (for all 10 embryos) or inter- (ASTEC-Pm8 versus the rest) embryonic matching cells.

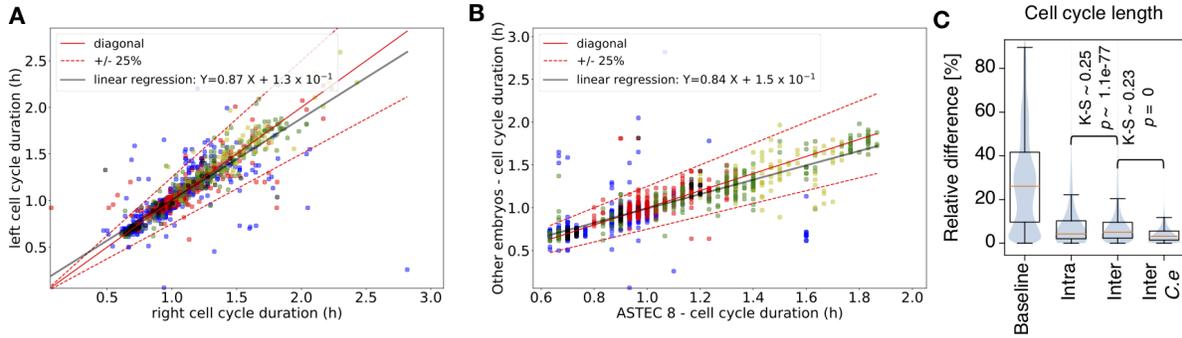


Figure S22: Cell cycle duration of matching cells. Cell cycle duration distribution (measured in HPF equivalent at 18°) of matching cells for A) left/right comparison in all embryos; B) pairwise comparison between ASTEC-Pm8 and each other embryo. Cells in each panel are color-coded with respect to their fate restriction: Epidermis (blue), Neural plate (red), Mesoderm (green), Endoderm (yellow) or still giving rise to several tissue fates (black). C) Violin plots showing the distributions of the relative differences in cell cycle length between all cells of all embryos (baseline), matching bilateral cells within a *Phallusia* embryo (intra), homologous cells across *Phallusia* embryos (inter) and homologous cells across 28 *C. elegans* embryos (Inter C.e.). Boxes show the first, second and third quartiles, whiskers the range to 1.5 interquartile.

Tree-edit distance. Ascidian cell lineage trees are expected to be bilaterally symmetrical (see for example Fig. S21). To compare the architecture of two cell lineage trees and quantify their structural similarity, we used a tree-edit distance approach based on the definition of a set of atomic edit operations applicable to tree structures (64). These edit operations typically consist of adding, deleting or substituting a node in trees. To compare two tree structures, say T_1 and T_2 , the minimal number of edit operations necessary to transform T_1 into T_2 is computed (Fig. S23), then normalized by the total number of nodes in both compared trees, which makes it possible to compare distances among trees of varying sizes.

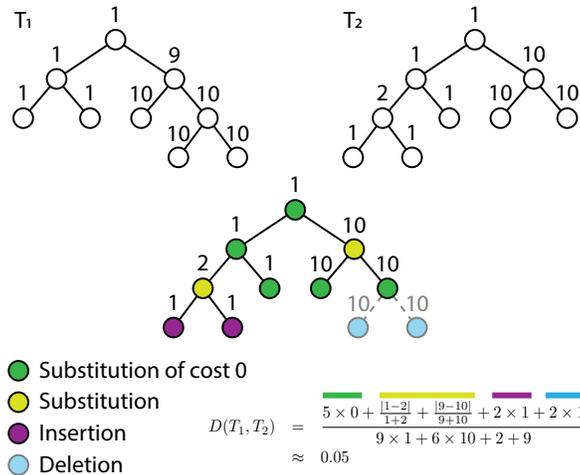


Figure S23: Principle of tree-edit distance computation between lineages. Two lineage trees T_1 and T_2 are compared. The numbers above the nodes represent the time associated with these nodes. T_1 is rewritten as T_2 . The distance between these two trees is 0.05 (5% of mismatch).

Of the many variants of tree-edit metrics defined in the literature (64), we opted for a topological metric augmented with the cell cycle length (i.e. the node structure of the compared trees is considered together with the length of the cell cycle), and considered that the trees are unordered (i.e. no ordering distinction is made among the children nodes of a node). This tree-edit distance (65, 66), is based on the definition of a set of elementary operations (such as replace, insert, delete) acting on the cells of the lineage trees, and for which a cost has been defined (see below). The cost of the sequence of edit operations with minimal cost, and normalized by the size of the trees, defines the distance between the compared lineage trees. The distance is a real between 0 and 1, and can be interpreted as an average cost of transformation per cell, hence reflecting the similarity between the two trees independently of their size (0 meaning that the two trees are exactly the same).

More precisely, (65) introduced an efficient algorithm to compute the minimal number of editions between two trees (in the case of binary trees, which is our case, computational time is essentially proportional to the product of the size of the compared trees), which was previously adapted to a biological context and used to compare the architecture of land plants (66). In the context of this metric, we used lineage trees where a single node is a cell throughout its period of existence, its cell cycle duration being an attribute of this node (n being a node, $t(n)$ is the value of the time attribute of this node). Substitution, insertion and deletion costs have to be provided. We set the insertion and deletion costs to 1. The substitution costs has to be strictly smaller than the cost of an insertion followed by an deletion (or the other way around). We therefore chose to set the substitution cost between two nodes n_1 and n_2 , $d(n_1, n_2)$ as the absolute value of the time difference between these two nodes normalized by the sum of the time of these two nodes:

$$d(n_1, n_2) = \frac{|t(n_1) - t(n_2)|}{t(n_1) + t(n_2)}. \quad (7.2)$$

By definition, the cumulative distance between is the cost of the transformation with minimal cost that transforms T_1 into T_2 using a series of edit operations as defined above (see Fig. S23). This distance is then normalized by the combined size of the trees. The final distance $D(T_1, T_2)$ can therefore be interpreted as an average distance per node corresponding to this optimal transformation.

To further assess the stereotypy of *Phallusia* development, we used the tree-edit distance to compare the structure of cell lineages originating from equivalent cells within embryos (by comparing left to right cells). We also compared the lineages between embryos by comparing cells that share the same name regardless of the left/right symmetry (ie A7.1 in a first embryo will be compared to A7.1 and A7.1* in the second embryo). In order to make the comparison between embryos meaningful, the developmental rates of embryos were normalized using the method described above prior to the computation of distances.

7.2 Spatial stereotypy

In order to assess the level of spatial stereotypy in *Phallusia mammillata* developing embryos, we employed several metrics to compare different properties of either *homologous* (for inter-embryonic analyzes) or *bilaterally symmetric* (for intra-embryonic analyzes) cells, collectively referred to as *matching* cells in what follows. Cell names were used to identify homologous or bilaterally symmetrical cells.

Moreover, we built a set of cell pairs that could be analyzed in the ten digitized embryos, this set of cells will be referred too as \mathcal{C}_p for the remaining of the text. To ensure that enough information was available to measure the different metrics (especially the lineage distance), we selected cells, among these

ten embryos, such that at least one of the two bilateral cells or one of the two bilateral sister cells of this cell divides twice in at least once for one of the ten embryos.

Ascidian embryos develop with bilateral symmetry. In agreement, tissue progenitors/anlage were bilaterally symmetrical with respect to the sagittal plane of the embryo (Fig. S16). Bilaterality of position was also observed for corresponding left and right cells sharing the same name (not shown).

7.2.1 Cell volume comparisons

To confirm the precision of our volume estimations, we first compared the sum of the volume of daughters to that of their mother in ASTEC-Pm1. The distribution of progeny/mother volume ratios shows that in more than 93% of cell divisions, the combined volume of the two daughter cells is within 5% of that of their mother (Fig. S24).

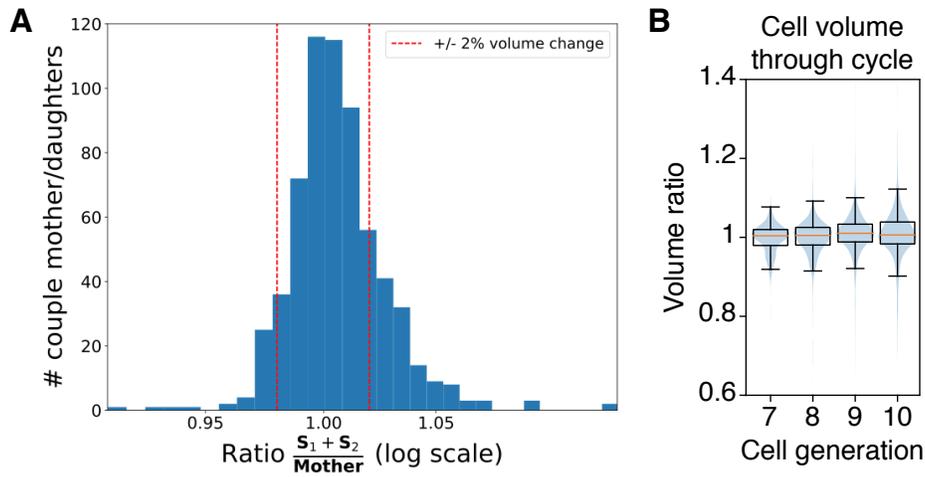


Figure S24: Analysis of cell volumes across cell division in ASTEC-Pm1. A) Distribution of the ratio of the sum of the volumes of the two daughter cells just after their birth over that of their mother just before its division. B) Cell volume ratio between the beginning and the end of the cell cycle, showing very little, if any, volume change throughout the cell cycle. Boxes show the first, second and third quartiles, whiskers the range to 1.5 interquartiles.

Next, we compared the volumes of matching cells within and across embryos. Because of a high variability in the total volume of individual embryos, we first normalized the volume of individual cells. Be A and B two matching cells, and be V_A and V_B their average volumes in μm^3 . Be also V_A^E (respectively V_B^E) the total volume of the embryo containing cell A (respectively cell B), which is almost perfectly conserved in time (not shown), and the fractional (normalized) volumes $\tilde{V}_i = \frac{V_i}{V_i^E}$, $i = A, B$. The volume comparison metrics \mathcal{M}_V is thus defined as the relative variation of normalised volumes between matching cells and given as

$$\mathcal{M}_V(A, B) = \frac{|\tilde{V}_A - \tilde{V}_B|}{\tilde{V}_A + \tilde{V}_B}, \quad (7.3)$$

which has by construction values in the interval $[0, 1)$, being 0 for perfectly equal normalised volumes and approaching 1 for strongly different \tilde{V}_i .

Fig. S25 presents a quantification of the degree of stereotypy in the volumes of intra- or inter-embryo matching cells. Specifically, Fig. S25A shows the volume distribution for each pair of bilaterally symmetric cells in all 10 embryos analyzed, while Fig. S25B shows volume distribution of homologous cells between ASTEC-Pm8 and the other 9 embryos.

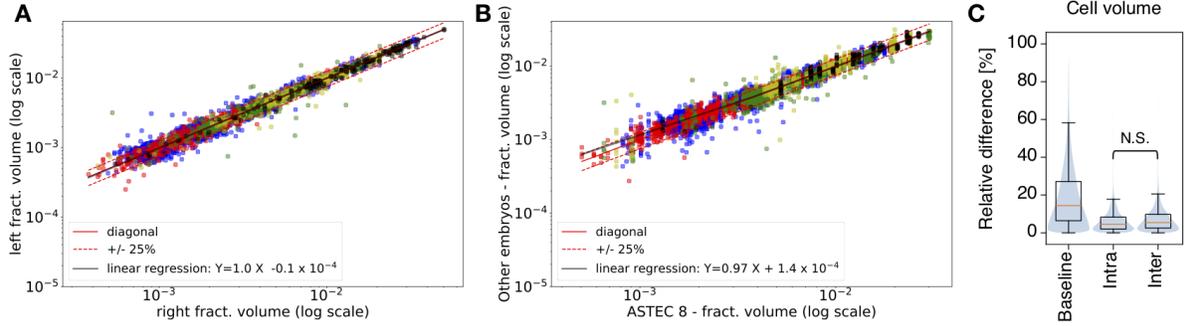


Figure S25: Volume of matching cells. Relative volume distribution (logarithmic scale) of matching cells for A) left/right comparison in all embryos; B) pairwise comparison between ASTEC-Pm8 and each other embryo. Cells in each panel are color-coded with respect to their fate restriction: Epidermis (blue), Neural plate (red), Mesoderm (green), Endoderm (yellow) or yet unrestricted (black). C) Violin plots showing the distributions of the relative differences in cell volumes between matching bilateral cells within a *Phallusia* embryo (intra) or homologous cells across *Phallusia* (inter). Boxes show the first, second and third quartiles, whiskers the range to 1.5 interquartile.

7.2.2 Analysis of relative positional variability

To quantify the relative positional variability of homologous cells between embryos (PV), we used the metric described in (23). This metric $PV(c_i, c_j)$ can be computed for every pair of cells c_i and c_j that coexist during a developmental interval between two embryos and is defined as follows:

$$PV(c_1, c_2) = \frac{1}{|T|} \sum_{t \in T} \sqrt{\sum_{n \in \mathcal{N}_t} \frac{(|\overline{c_{1t}}\vec{n}|_2 - |\overline{c_{2t}}\vec{n}|_2)^2}{|\mathcal{N}_t|}} \quad (7.4)$$

where c_1 and c_2 are cells from two different embryos (note that the cells being from different embryos is not a mathematical requirement), $\overline{c_{it}}\vec{n}$ is the vector between the position of the barycenters of the cells c_i and n at time t , $|\bullet|_2$ is the L2 norm, T is a discrete set of percentages of cell cycle for the two cells c_1 and c_2 (for example, $\overline{c_{i5}}\vec{n}$ is the vector between the cells c_i and n at 5% of the cell cycle of c_i , in our case we used $T = \{0, 5, 10, \dots, 100\}$), \mathcal{N}_t is the set of cells that are present in both embryos at time t . We re-implemented the metric as described in (23) with a minor change: instead of doing a nearest neighbour interpolation in time between embryos, we did a piece-wise linear interpolation, which allowed to look at precise windows of the cell cycle (Fig. S26 shows the computed positional variability mapped on the lineage tree).

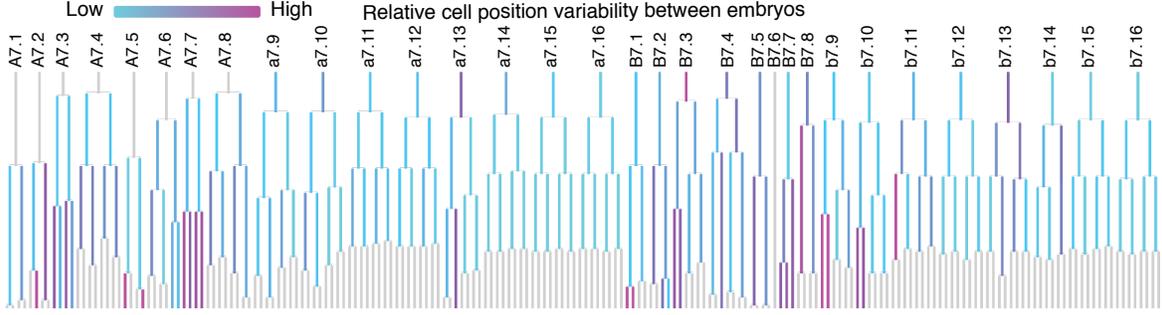


Figure S26: Cell position variability. Projection onto the cell lineage tree of the median level of variability in cell positions across all 10 segmented embryos. Cells for which this value cannot be computed are represented in grey. This panel was built using Tulip (63).

7.2.3 Cell neighbourhood comparisons

In order to quantify the similarity of neighbourhood between two matching cells, one must keep in mind that each cell can establish and lose contacts during its life. Some of these contacts will be transient, lasting for a few minutes only, and others will only involve a small portion of the cell surface. We define our neighbourhood comparison metrics for a pair of matching cells A and B (see Eq. (7.6) below) as the highest fraction of common neighbours at any two possible time points of existence of A and B. In order to get rid of “noisy” contacts, we impose at each timepoint a threshold on the contact surface, such that neighbours having a contact with A smaller than 5% of the surface of A are discarded.

One can then define, for each cell A, the evolution of its neighbours in time, which is a function $\mathcal{N}_A(t)$ associating with each time t the list of names of its K neighbours at this time: $\mathcal{N}_A(t) = \{n_1^A, n_2^A, \dots, n_{K(t)}^A\}$. Given the discreteness of our dataset in time, the function $\mathcal{N}_A(t)$ reduces to a collection of neighbours lists, one for each timepoint of existence of the cell A.

Consider now a pair of matching cells A and B, and be $\mathcal{N}_A(t)$ and $\mathcal{N}_B(t)$ their neighbours evolution functions. These two matching cells may exist at different developmental stages (although this event is extremely rare), such that these two functions might also be defined on fully disjoint time intervals. We refer to the set of time points of existence of cell A (respectively B) as $S_t(A)$ (resp. $S_t(B)$). For each pair of time points $t_i^A \in S_t(A)$, $t_j^B \in S_t(B)$, we define the instantaneous neighbours similarity score as

$$\mathcal{K}_N(A, B, t_i^A, t_j^B) = 2 \frac{|\mathcal{N}_A(t_i^A) \cap \mathcal{N}_B^\dagger(t_j^B)|}{|\mathcal{N}_A(t_i^A) \cup \mathcal{N}_B^\dagger(t_j^B)|}, \quad (7.5)$$

where $|\cdot|$ stands for set cardinality and \cap and \cup are here the set intersection and union operations. In Eq. (7.5), $\mathcal{N}_B^\dagger(t_j^B) = \mathcal{N}_B(t_j^B)$ if B is the homologous of A and $\mathcal{N}_B^\dagger(t_j^B) = \{\mathfrak{s}(n_1^B), \mathfrak{s}(n_2^B), \dots, \mathfrak{s}(n_{K(t)}^B)\}$ if B is the symmetric of A, where $\mathfrak{s}(c)$ stands for the name of the bilaterally symmetric cell of cell c.

The neighbours comparison metrics is then defined as

$$\mathcal{M}_N(A, B) = \max \{\mathcal{K}_N(A, B, t_i^A, t_j^B)\}_{t_i^A \in S_t(A), t_j^B \in S_t(B)}, \quad (7.6)$$

i.e., it is the highest fraction of common neighbours between any two possible instantaneous neighbours lists of A and B. This metrics gives values in the interval $[0, 1]$: a value of 1 means the two cells have exactly the same neighbourhood (up to contacts that are discarded as noisy), while a value of 0 means no common neighbour is present at any time.

Fig. S27A shows that common neighbours between homologous cells among all 10 embryos (blue histogram) are largely more stable in time than the non-common ones (red histogram). Fig. S27C shows the evolution of contact variation/conservation across cell cycles 7, 8, 9 and 10. The contact duration is expressed, for each pair of matching cells, in units of the average of their cell cycle durations. In agreement with the long duration of common cell-cell contacts, most cells kept the same neighbours throughout their life (Fig. S27B). In other terms, there was no individual cell migration, and only extremely rare cell rearrangements involving the loss of at least 1/4 of the neighbours between the birth and division of the cell.

7.2.4 Comparison of areas of cell-cell contact

The metrics $\mathcal{M}_N(A, B)$ defined above provides information about the identity of conserved neighbours, but does not describe the evolution of the size of the contacts conserved between matching cells. When comparing the size of contacts rather than the identity of neighbours, one must note that more than one pair of time points (and hence of instantaneous neighbourhoods) can maximize Eq. (7.6). This yields therefore possibly different lists of neighbours, which may be characterized by largely different contact surfaces.

To overcome this problem, given two matching cells A and B, their neighbours evolution functions $\mathcal{N}_A(t)$ and $\mathcal{N}_B(t)$ and their neighbours comparison score $\mathcal{M}_N(A, B)$, we select all possible pairs of time points $t_i^A \in S_i(A), t_j^B \in S_i(B)$ such that the instantaneous neighbours similarity score $\mathcal{K}_N(A, B, t_i^A, t_j^B)$ equals the neighbours comparison score $\mathcal{M}_N(A, B)$. Be $T_{A,B} = \{(t_{i_1}^A, t_{j_1}^B), (t_{i_2}^A, t_{j_2}^B), \dots, (t_{i_W}^A, t_{j_W}^B)\}$ the set of the W selected pairs. For each pair $(t_i^A, t_j^B) \in T_{A,B}$ we define the instantaneous contacts comparison score as

$$\mathcal{H}_C(A, B, t_i^A, t_j^B) = \left\langle \frac{|\tilde{C}_k^A - \tilde{C}_k^B|}{\tilde{C}_k^A + \tilde{C}_k^B} \right\rangle_{\text{CN}}, \quad (7.7)$$

where $\langle \cdot \rangle_{\text{CN}}$ is the average value over all common neighbours k of A and B and we introduced the normalized contact surfaces $\tilde{C}_k^X = \frac{C_k^X}{S_X}$, $X = A, B$, where C_k^X is the contact surface in μm^2 between cell X and neighbour k and S_X is the total surface of X in μm^2 . Hence eq. (7.7) is the average value of the relative change in normalized contact surfaces with instantaneous common neighbours.

We define the contact comparison metrics as

$$\mathcal{M}_C(A, B) = \min \{ \mathcal{H}_C(A, B, t_i^A, t_j^B) \}_{(t_i^A, t_j^B) \in T_{A,B}}. \quad (7.8)$$

Equation (7.8) is the minimal average relative variation of contacts with conserved neighbours, with values in $[0, 1]$: it is 0 if each normalised contact with common neighbours is exactly the same for the two matching cells A and B and tends to 1 if all contacts are markedly different. Fig. 2D shows the evolution of this metric over time for intra- and inter-embryonic matching cells and Fig. S27C shows the evolution of this metric over time for epidermal and non-epidermal cells.

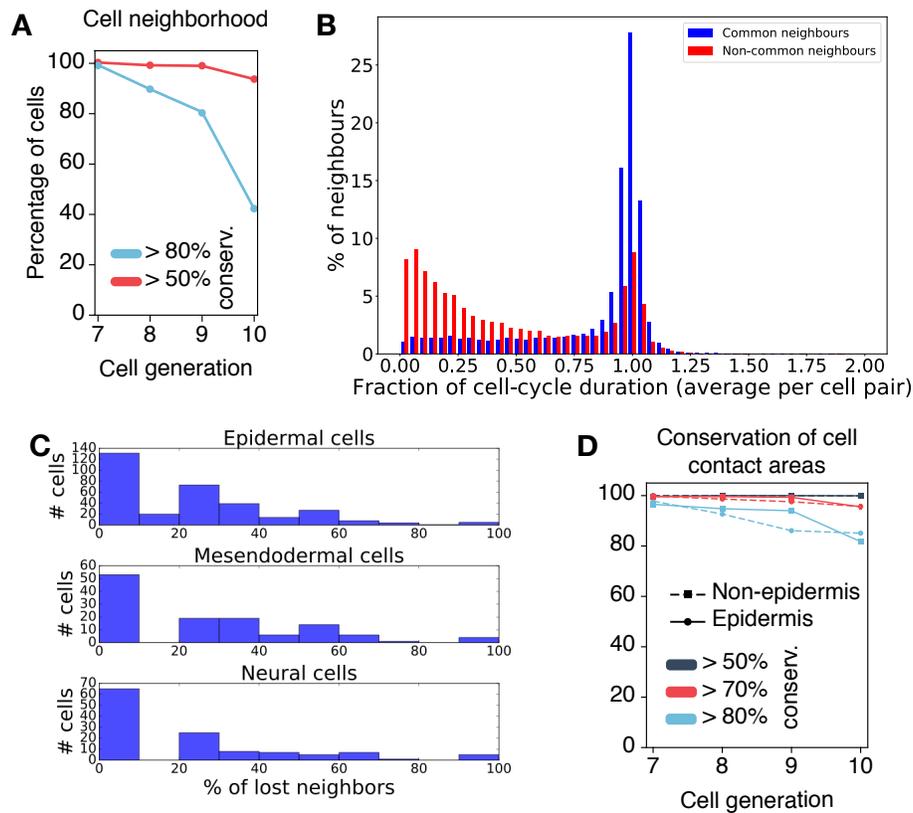


Figure S27: Duration and stability of cell-cell contacts. A) Evolution in time (for cell generations 7 to 10) of the percentage of matching cells showing a conservation in neighbourhood larger than the indicated thresholds. B) Histogram for the distribution of contact duration for common neighbours (blue) and non-common neighbours (red) of matching cells between each possible pair of ASTEC-Pm1 to -Pm10 embryos (cell generations 7 to 10). The duration is measured in units of the average cell cycle duration of each pair of matching cells. C) Stability of contacts established by a cell with other cells within each tissue in ASTEC-Pm1 (cell generations 7 to 10): percentage of cell contacts within the epidermis (top), the mesendoderm (middle) and the neurectoderm (bottom) lost over the entire life of the cell. D) Evolution in time (for cell generations 7 to 10) of the percentage of matching cells showing a conservation in contact area to common neighbours larger than the indicated thresholds within epidermal cells (solid lines) or within non-epidermal cells (dashed lines).

7.2.5 Comparative analysis of the degree of stereotypy in *P. mammillata* and *C. elegans*

To relate the level of stereotypy in *P. mammillata* to that of another model organism, we computed the same descriptors in *C. elegans* from the 28 segmented and lineaged embryos provided in (23). As *C. elegans* segmentation and tracking was done with nuclei rather than whole cells, the volume stereotypy analysis was not possible for *C. elegans*. The other metrics were computed as in *P. mammillata* and, in the case of the lineage tree edit distance, since its value can depend on the depth of the lineage trees, the *C. elegans* lineage trees were cropped so they start and end with similar number of cells as the *P. mammillata* lineage trees (i.e. starting with ~ 64 cells and ending with ~ 372 cells).

7.2.6 Comparison between distributions

To test whether distributions were significantly different we used the non-parametric Kolmogorov-Smirnov test (67) since we could not make assumption about the distributions to test (normal, gamma or other type of distribution).

8 Analysis of fate restriction events during ascidian development (Supplement to Figure 3)

During development, embryonic cells progressively see their developmental potential restricted to a single tissue fate. This progressive differentiation of cells involves successive types of events. First, cells become fate-restricted within the context of the embryo: a cell division giving rise to two sister cells of distinct larval fates is referred to here as an asymmetric division and corresponds to a fate restriction event. At this stage of differentiation, the sister cells may still be epigenetically identical, their distinct fates resulting from their subsequent exposure to signals from differing cellular environments. Second, these sisters, or their progeny, adopt distinct epigenetic states and start expressing distinct sets of genes, a process called fate specification. At this stage, cells usually adopt their final fate even if they are explanted and placed in a neutral environment such as a petri dish. Fate acquisition can still be influenced by extracellular signals. Finally, when their fates become determined, the cells will irreversibly adopt their final fate even if they are placed in an ectopic cellular environment, independently of the extracellular signals emitted by this environment.

Irrespective of the fates of the daughter cells, a cell division can not follow Hertwig's rule (68), be geometrically unequal producing two daughters of different volumes, or give rise to two daughters with different cell cycle durations (i.e. which divide asynchronously).

8.1 Automatic detection of cell fate restriction events

To automatically detect cell fate restriction events, we observed that known cell fate restriction events usually give rise to sister cells exhibiting differences in their division patterns, their volumes, their cell cycle duration or a combination of these three parameters. We built upon these qualitative observations to design metrics predicting cell fate restriction events.

8.1.1 Detection of asymmetric cell divisions

First, we explored the relationships between cell lineage architecture and cell fate. Using the cell lineage tree-edit distance defined above applied to the 10 embryos and averaged across them, we found

that 64-cell stage embryonic cells fated to a given tissue give rise to more similar lineage trees than cells fated to distinct tissues (Fig. S28).

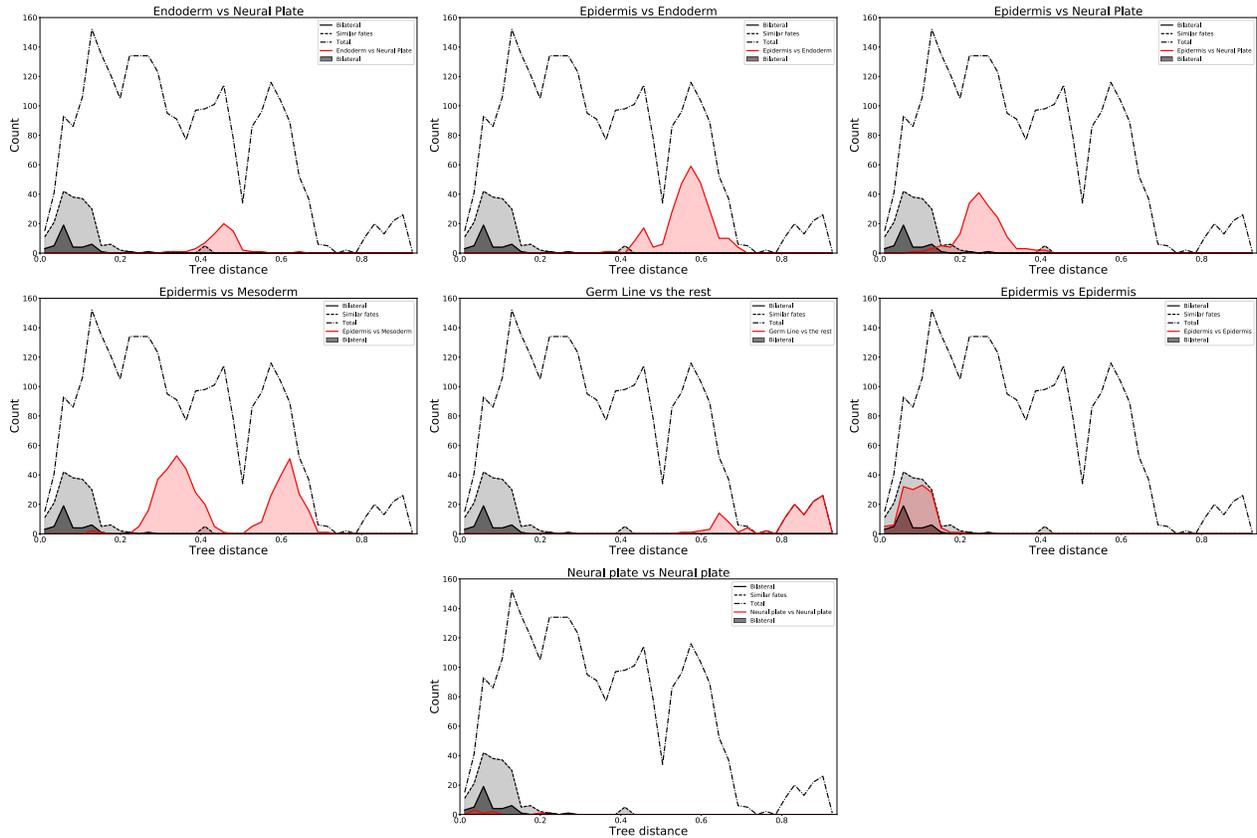


Figure S28: Distribution of the distances between cell lineage trees within embryos. Mixed hashed line: pairwise comparison between all trees originating from a single progenitor at the 64-cell stage. Hashed line: pairwise comparison restricted to the similarly fated cell pairs of the 64 cell-stage. Plain line: pairwise comparison restricted to bilaterally symmetrical cell pairs at the 64 cell-stage. Red line: pairwise comparison restricted to 64-cell stage cells restricted to the fates indicated in the insets. ASTEC-Pm1 to Pm10 datasets.

We then clustered all 64-cell stage cells, averaging across the 10 embryos, according to the pairwise tree-edit distance of the cell lineage trees they seeded. We used an ascending hierarchical clustering method, aggregating the distances with the Ward method (69). This showed that cell lineage similarity groups 64-cell stage cells according to their larval fates (Fig. S29), and therefore that mitotic histories are indicative of fates.

Having established this relationship between cell lineage architecture and cell fates, we next compared the cell lineage trees originating from sister cells after each cell division. To have sufficient cell lineage architecture information, this approach was carried out for bilaterally symmetrical couples of sister cells for which one of the sisters (among the four) divides at least twice until the end of the movie. We reasoned that sister cells giving rise to similar cell lineage trees probably adopt the same fate, while sister cells seeding markedly different lineage trees may adopt distinct larval fates.

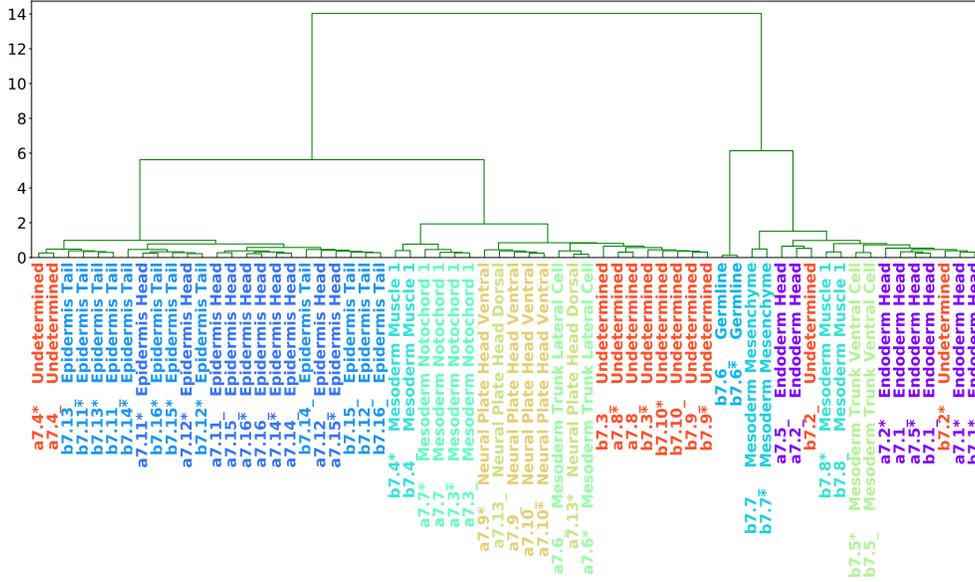


Figure S29: Hierarchical clustering of individual cells at the 64-cell stage, according to the tree-edit distance between the lineage trees they seed, within embryos. See text for details of clustering method. Colors represent larval fates. ASTEC-Pm1 to Pm10 datasets.

Sister cell lineage tree inequality index. Let $S1$ and $S2$ be two sister cells (and $S1'$ and $S2'$ their bilateral homologs), T_{S1} and T_{S2} be the lineage trees originating from $S1$ and $S2$ (resp. $T_{S1'}$ and $T_{S2'}$ originating from $S1'$ and $S2'$), then the cell lineage tree inequality index ($Lted_i\{S1, S2\}$) is the minimum of the lineage tree distance between the lineage trees originating $S1$ and $S2$ and the same distance between the lineage trees originating from $S1'$ and $S2'$:

$$Lted_i\{S1, S2\} = \min(D(T_{S1}, T_{S2}), D(T_{S1'}, T_{S2'})) \quad (8.1)$$

Using the set of cells C_p described in 7.2 we looked for cells that were consistently having an unequal division according to our indices Lt_i and V_i . To do so we had to set a threshold value above which we assumed that the cells were dividing unequally. We reasoned that if a cell is dividing unequally then the orientation of this inequality should always be towards the same direction (i.e. if daughter 1 is larger than daughter 2 then it should be the case for all embryos). To look at this we computed an unequal division agreement score accounting for the agreement of the largest (resp. longest lived) daughter for V_i (resp. Lt_i). This score goes from 50%: amongst the embryos, half of the largest (resp. longest lived) cell was D_1 and the other half was D_2 , to 100%: it was always the same daughter that was the largest (resp. longest lived). Using this metric, we set our threshold for the V_i (resp. Lt_i) value such that at least 90% of the cells were in agreement 100% of the time (see Fig. 3).

We then wanted to test whether the lineage tree edit distance was predictive of asymmetric divisions. To do so we split \mathcal{D} , the set of sister cell pairs, into three sub-sets: sister cell pairs known to give rise to distinct cell fates, and thus originate from the asymmetric division of their mother (see Table S4, n=28); sister cell pairs known to adopt the same cell fate and thus originate from the symmetric division of their mother (see Table S5, n=30) and the remaining sister cells, whose fate status was unknown (n = 46).

We then trained a Quadratic Discriminant Analysis classifier (70) on these sets of known symmetric and asymmetric divisions. We trained the model with every single cell pair and then predicted on the metrics averaged over the embryos, for each cell pair on the lineage tree edit distance index ($Lted_i$). Using that method with the sister cell lineage tree inequality index, we correctly classified 53 cell pairs over 58 (28 known asymmetric divisions and 30 known symmetric divisions). This procedure missed 5 asymmetric cell divisions (false negatives), and identified all symmetric divisions (no false positives). These results show that 1) if the model predicts an asymmetric division it is likely to be true (no false positives) and 2) that some asymmetric divisions occur without short- to mid-term differences in their cell lineages, and hence go undetected. Running this trained classifier on sister cell pairs of unknown status led to the identification of 12 candidate asymmetric cell divisions (see Table S6) and 34 symmetric divisions.

8.1.2 Detection of geometrically unequal cell division and differential cell cycle duration

To identify unequal cell divisions and divisions giving rise to sister cells with unequal cell cycle durations, we defined two metrics: the sister cell volume inequality index and the sister cell cell cycle duration inequality index.

Sister cell volume inequality index. Let $S1$ and $S2$ be two sister cells (and $S1'$ and $S2'$ their bilateral homologs), such that $Vol_{S1} < Vol_{S2}$ and $Vol_{S1'} < Vol_{S2'}$. The sister cell volume inequality index $V_i\{S1, S2\}$ was computed as follows:

$$V_i\{S1, S2\} = \min \left(\frac{Vol_{S1}}{Vol_{S2}}, \frac{Vol_{S1'}}{Vol_{S2'}} \right) \quad (8.2)$$

Sister cell cell cycle duration inequality index. Let $S1$ and $S2$ be two sister cells (and $S1'$ and $S2'$ their bilateral homologs), such that their cell cycle duration (CCL_{S_i}) follows $CCL_{S1} < CCL_{S2}$ and $CCL_{S1'} < CCL_{S2'}$. The sister cell cell cycle duration inequality index $Lt_i\{S1, S2\}$ was computed as follows:

$$Lt_i\{S1, S2\} = \min \left(\frac{CCL_{S1}}{CCL_{S2}}, \frac{CCL_{S1'}}{CCL_{S2'}} \right) \quad (8.3)$$

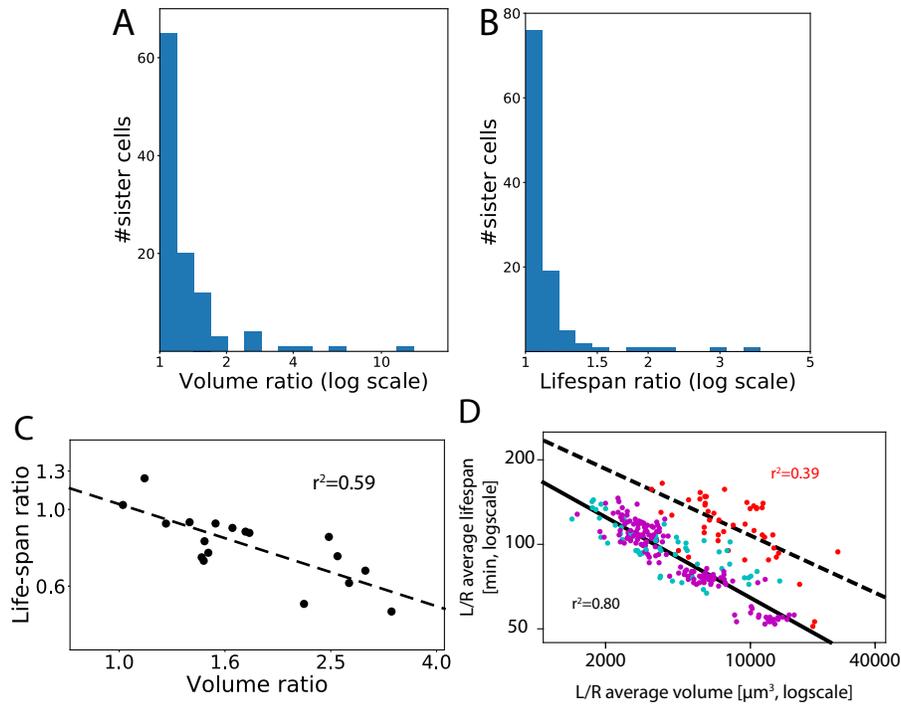


Figure S30: Sister cell volumes or cell cycle durations ratios in *ASTEC-Pm1*. A) Distribution of the sister cell volume ratios. B) Distribution of the sister cell cell cycle duration ratios. C) Relationships between sister volume ratio and cell cycle duration ratios in neural plate cells generated by the division of stage 12 neural plate precursors in Fig. 3D-F. D) (Logarithmic scale) Relationships between volume and cell cycle duration for epidermis (magenta), neural plate (cyan) and endoderm (red). Power-law regressions are calculated for epidermis+neural plate (black) and mesoderm (red).

Fig. S30A and B) presents an analysis of the distributions of the sister cell volume inequality index and the sister cell cell cycle duration inequality index.

8.1.3 Clone shape analysis

In order to look at division orientation we looked at the shape of clones of 4 cells issued of the same grand-mother cell. The topology of the contacts between the cells informs shape of the clone (see Fig. S31 A).

9 Description of the differential sister cell induction model (Supplement to Figures 4 and 5A)

9.1 Model summary

In ascidians, most cell fate decision events are due to cell-cell communication events called embryonic inductions (11). Previous work suggested that in early ascidian embryos, some inductions act at very short range, and that the area of contact between cells emitting and responding to an inducer provides the quantitative information that decides on the outcome of the induction (12, 32).

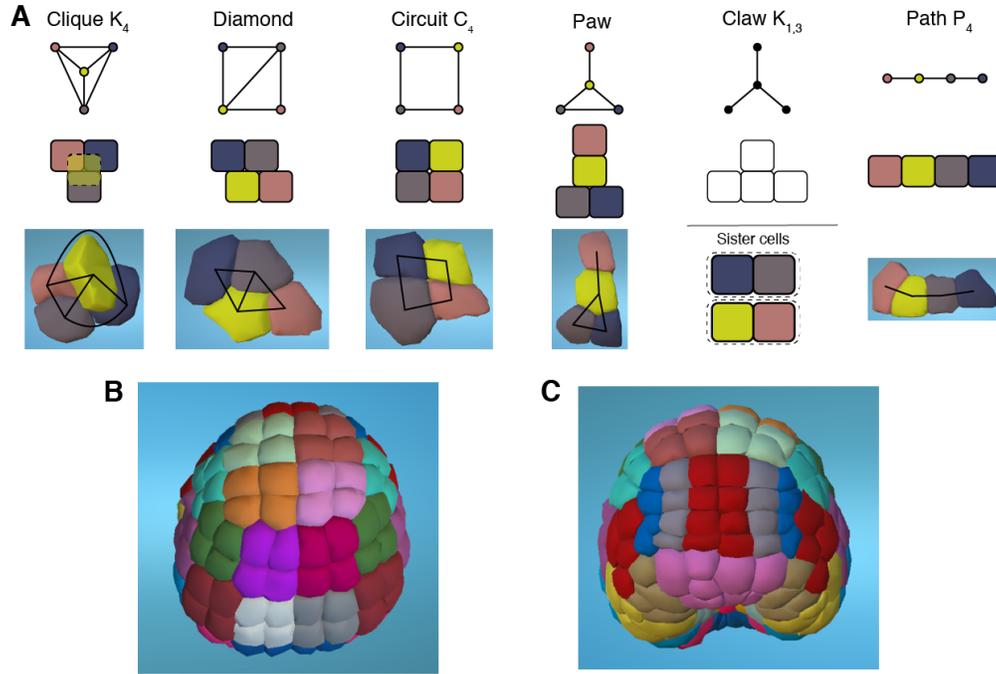


Figure S31: Shapes of 4-cell clones. A) Enumeration of all the possible connected graphs with 4 vertices, their corresponding theoretical cell configuration, classification between square, T/triangular or linear shape and corresponding examples in the embryo. Dark blue and marron cells are sister cells, yellow and pink cells are sisters. In parenthesis are the number of clones following the configuration. B) A ventral view of the geometry of 4-cell epidermal clones at the late gastrula stage (Stage 12). Note that all clones have a square shape. C) A dorsal view of the neural plate and adjoining epidermal cells. Note that all central 6 a-line neural plate clones have a linear shape. In B and C, each color represents a clone. Bilaterally homologous clones usually have the same color, except for homologous midline clones whose color can differ.

We developed a computational model to test the generality of this induction mode by integrating the geometrical cell-cell contact information from our embryos with an atlas of spatio-temporal gene expression for extracellular signaling ligands and inhibitors (71, 72), which was extracted from the ANISEED database (61).

Here is a summary of the model, in-depth description is available in the next subsections:

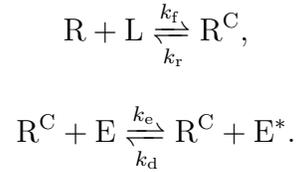
First, we translated gene expression patterns extracted from the ANISEED database (61) into protein production maps and projected it onto our segmentation, hence combining gene expression patterns and cell geometry. We examined the six major pathways involved in early ascidian development: FGF, EphrinA, Notch, Bmp, Nodal, and Wnt (see see 9.2).

The following assumptions were taken to simplify the model (see 9.3):

1. Cell fate restriction events in ascidians are driven by short-range cell inductions.
2. All cells are equally competent to respond to extracellular signals emitted by their physical neighbors

3. The range of extracellular signaling proteins (ligands and antagonists) is either first-order neighbors (scenario 1) or second-order neighbors (scenario 2).
4. The action of sequestering signaling antagonists is dominant over signaling ligands at interfaces where both are present.
5. Ligand concentrations at cell-cell interfaces, receptor densities and ligand-receptor association/dissociation constants are considered equal for all signaling pathways.

We computed pathway activation based on the quantitative geometry of cell contacts and qualitative expression profiles using the law of mass action and the derived following equations:



Where R is the number of receptors, L the number of ligands, R^C the number of complexed receptors, E the number of effectors, E^* the number of complexed effectors and k_r, k_f, k_d, k_e the reaction constants (see 9.4, 9.5). In the case of the antagonistic action of FGF and Eph signaling in the ERK pathway, we used slightly different equations (see 9.6).

To introduce the geometry of cell contacts into these equations, and according to the conservation of mass, the total (complexed and free) receptor concentrations $R^T(t)$, satisfy:

$$R^T(t) = R(t) + R^C(t), \quad (9.1)$$

with:

$$R^T(t) = \alpha_R \cdot A(t). \quad (9.2)$$

where $A(t)$ is the surface of the cell (μm^2) exposed to free ligand and α_R (*receptors*. μm^{-2}) is the local receptor density at the surface of a cell, considered equal for all cells and uniform on the surface of each cell.

The previous equations, when integrated, output the concentration of activated effectors at the membrane for every cells and for all the pathways considered. To convert continuous levels of effector activation into a binary map of induced/uninduced cells, we implemented induction thresholds under which there is no induction and above which there is induction, governed by tunable free parameters and defining both minimum effector activity levels necessary for an induction and a minimum ratio of activity between induced and non-induced cells (see 9.7).

Our simplified differential induction model thus provides both a semi-quantitative estimate of the relative level of signaling for each pathway in each cell and a qualitative prediction of differential inductions. To do so, we used a set of cells with known inductions and a set of cells with known absence of induction.

A small number of free parameters describe our model (Table S12):

- $k_f.L_T$ [s^{-1}]: the association constant of complexed effectors (k_f) times the total concentration of ligands at the membrane (L_T)
- $\alpha_R.k_e$ [$\mu\text{m}^{-2}s^{-1}$], where α_R is the receptor density and k_e is the association constant between an effector and a complexed receptor
- $\delta_1 = \frac{k_n}{k_m}$, where k_n is the association constant between RasGTP and activated RasGAP and k_m is the association constant between RasGDP and activated SOS
- $\delta_2 = \frac{k_p.X^*}{k_m}$ [$\#.\mu\text{m}^{-3}$], where X^* is an unknown protein acting on RasGTP, k_p is the association constant between RasGTP and X^*
- Time of induction [s]: the time necessary for a cell to integrate a membrane signal.

To convert continuous values of effector activation into binary inductions, we introduced two tunable thresholds:

- ρ_0 : the threshold value for the ratio between the concentration of effectors between two cells, above which two sister cells are considered differentially induced
- τ : the maximum percentage of cells that can respond to exposure at a given pathway (i.e. be induced).

To determine the values of these parameter that produce model predictions that best fit biological knowledge, we performed a grid search across our parameters and compared the results of the model for each parameter combination to an experimentally-determined ground truth of differential inductions extracted from the literature (see 9.8 and Table S11).

The model was run on different embryos, under different assumptions, and in different experimental conditions (see 9.9 and 9.10). These different runs did not all involve the retraining of the model. We used the following logic to decide which set of parameters to use in each case (summarized in Table S13):

- To compare different model assumptions on the same embryo (e.g. range of inducers; area dependency), the model prediction was the best fit to experimental data obtained by exploring parameter space under the given assumptions. The optimal parameter set was therefore recomputed for each assumption. We chose this strategy because this is the most stringent one and it defines an upper bound.
- To assess the effect of genetic or geometric perturbations (e. g. Comparing Eph signaling inhibition to WT, or half embryo to WT or analysing the sensitivity of induction to changes in surfaces of contact) we did not recompute the optimal set of parameters. We applied the WT parameter set to the new experimental situation. This procedure allowed to focus on the effect of the experimental perturbation, all other things being equal.
- Finally, to study the robustness of model predictions to natural variability, we retrained the model for each embryo and compared the parameters obtained (Table S12). Retraining for each embryo was necessary since embryos may differ both in their size and in their genetic material, which can cause differences in levels of expression of proteins or changes in their association constants. Retraining the model allowed taking this into account.

9.2 An atlas of spatially-controlled signaling pathways during ascidian early embryogenesis

The concentration of individual secreted proteins and receptors are not known in ascidians. To identify potential inducers, we searched for signaling pathways with zygotically activated, and spatially restricted, ligands during early ascidian embryogenesis, based on *in situ* hybridization data in *Ciona intestinalis*. Gene expression profiles with cellular resolution for all ligands and inhibitors of signaling pathways were extracted from the ANISEED database (<https://www.aniseed.cnrs.fr/>) (61, 73). Based on the phylogenetic proximity of *Phallusia* and *Ciona*, and on published analysis of the conservation of orthologous gene expression patterns in these two species (28, 29) then considered that RNA expression profiles foreshadow protein expression profiles, with a protein synthesis/processing delay that varied between pathways and was estimated from biological data (Fig. S32). We also considered that the expression profiles of orthologous *Phallusia mammillata* and *Ciona robusta* genes are conserved and used the extensive *Ciona robusta* expression atlas as a proxy for *Phallusia mammillata* genes.

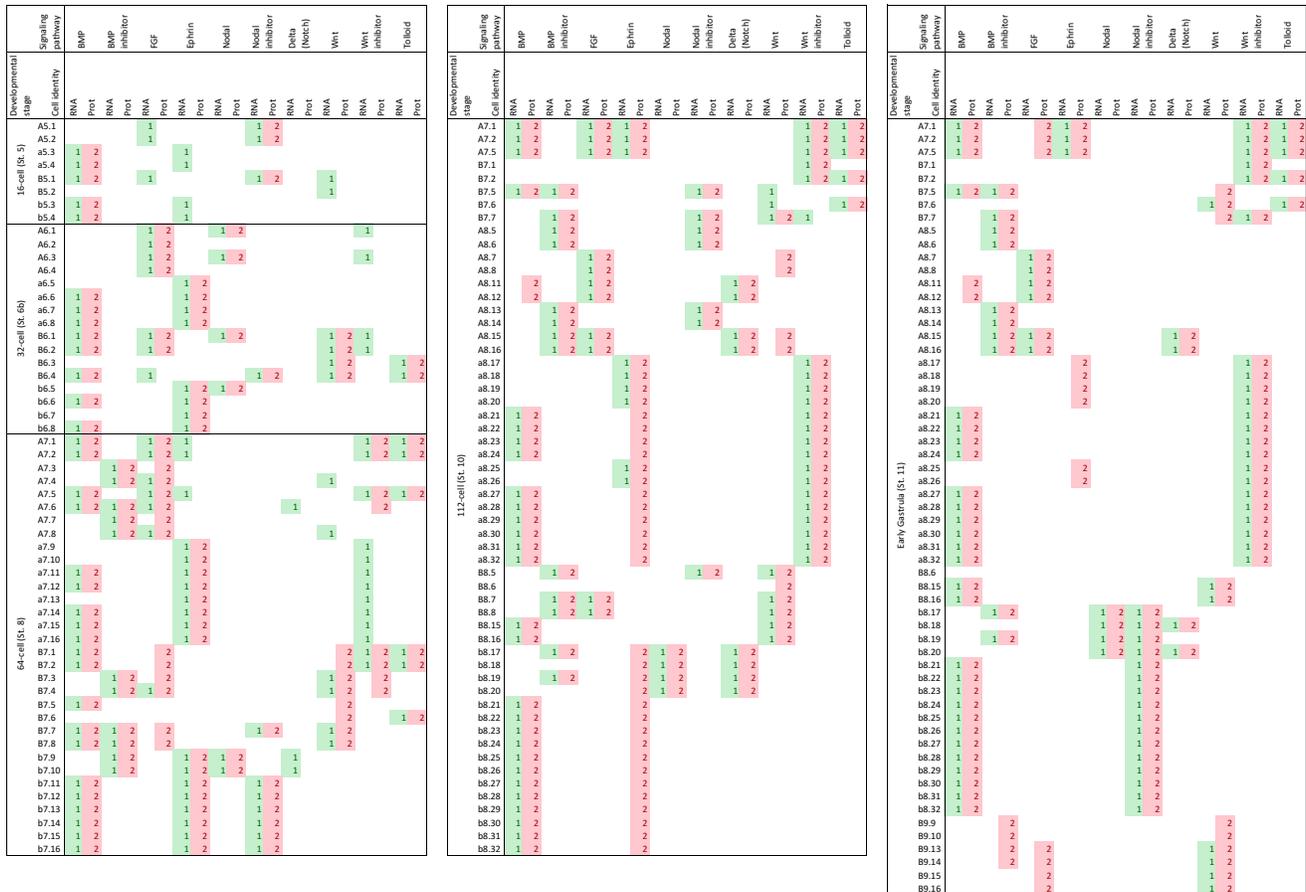


Figure S32: Pattern of expression of RNA and protein for the major signaling molecules up to the early gastrula stage in *Ciona robusta*. The Table shows aggregated data reflecting the expression at 5 developmental stages of ligands and antagonists for the indicated signaling pathways. Green (1): RNA expression, red (2): protein expression.

9.2.1 Signaling pathways with spatially-restricted gene expression profiles

Extracellular ligands and antagonists for only 6 major pathways previously involved in cell fate decisions are zygotically activated in a spatially-restricted manner up to the gastrula stage: FGF, EphrinA, Notch, Bmp, Nodal, and Wnt. Secreted molecules with a ubiquitous or maternal expression pattern were not considered. Aggregated ligand expression patterns for each pathway were obtained by operating a union of the expression patterns of the following individual ligands: Nodal pathway: Nodal (KH.L106.16); Wnt: Wnt unassigned 2 (KH.C9.257), Wnt pathway 10 (KH.C9.27), Wnt 5b (KH.L152.45); Notch pathway: Delta2 (KH.L50.6); BMP pathway: ADMP (KH.C2.421), derrière-like (KH.C4.547); FGF pathway: FGF9/16/20 (KH.C2.125), FGF8/17/18 (KH.C5.5) ; Ephrin pathway: EphrinAa (KH.C3.762), EphrinAc (KH.C3.52), EphrinAd (KH.C3.716).

The Nodal, Bmp and Wnt pathways are antagonized by the secreted inhibitors Lefty, Chordin/Noggin/Pinhead, and sFRP/Dkk, respectively, which act by sequestering the ligand and preventing their interactions with their receptors. Aggregated patterns of expression of secreted inhibitors for each pathway were obtained by operating a union of their expression patterns. Nodal: Lefty (KH.C3.411); Wnt: Sfrp2 (KH.C2.469), Sfrp1/5 (KH.L171.5), Dkk1/2/4 (KH.L20.29); Notch: no secreted antagonist expressed at the stages considered; BMP: Chordin (KH.C6.145), Pinhead (KH.C2.1069). Chordin can in turn be inactivated by the astacin-like zinc-dependent metalloprotease Tolloid (KH.C12.156). In the model, we consider that if Tolloid is present at a cell-cell interface, it fully inactivates Chordin at this interface.

Bmp3 (KH.C12.491), Bmp2/4 (KH.C4.125) and Noggin (KH.C12.562) were not taken into account in the model, because their expression level is much lower than that of other secreted signaling molecules (J. Piette, C. Dantec, unpublished observations, see the RNA-seq tracks of the *Ciona intestinalis* ANISEED genome browser), and because interference with the function of these ligands with antisense morpholinos does not cause significant phenotypes (72).

The action of the FGF pathway, which activates the small G protein *Ras* via the ubiquitous FGF receptor (KH.S742.2), is antagonized by the EphrinA signaling pathway acting via the Eph3 receptor (KH.C7.568) to repress the activation of *Ras* via the action of RasGAP (KH.L152.46) (74). In the absence of demonstrated evidence, we neglected the role of EphrinA reverse signaling during early ascidian embryogenesis.

Ascidian genomes are unduplicated (75), and for all 6 signaling pathways, a single receptor complex triggers the activation of the intracellular pathway (e.g. Alk4/5/7 (KH.L22.40) is the sole ascidian ortholog of vertebrates Alk4, Alk5 and Alk7). All developmental signaling pathway receptors considered here (FGFR: KH.S742.2; Ephrin: KH.C7.568; Notch : KH.C9.176; Nodal: KH.L22.40, KH.C1.598; Bmp: KH.C14.43, KH.C1.598; Wnt: KH.C9.236, KH.C9.260, KH.L9.43) are expressed maternally, and we therefore assume that they are expressed at the same level in all cells up to the early gastrula stage, around 5 hours after fertilization.

9.2.2 Estimation of spatio-temporal protein expression

From the RNA expression profile, we inferred the time needed for the processing, maturation and secretion of the signaling proteins from existing information on the time delay between their transcriptional activation and that of their earliest known target genes.

Nodal processing and signaling was considered rapid, occurring within one developmental stage, as its earliest target gene, *Delta2*, is expressed in the b6.5 lineage within 30 minutes of the onset of *Nodal* expression in the same lineage at the 32-cell stage (76). *Delta2* was also considered to be acting within 30 minutes of its transcriptional activation, the time delay needed for the activation of its first target gene, *Hes-b* in the secondary notochord lineage (76). Bmp transcriptionally activates *Chordin* in *Halocynthia* within a few minutes (77) and was also considered to act rapidly. By contrast, FGF was only considered to produce sufficient protein for induction 40 minutes, or one cell stage, after its transcriptional activation, as the onset of expression of FGF9/16/20 in vegetal cells is at the 16-cell stage but the ligand only activates ERK, its intracellular effector at the late 32-cell stage (78). Wnt secretion pathway is complex, involving numerous post-translational modifications (79) and, in the absence of data on Wnt target genes in ascidians, its ligand availability was considered to be slow and delayed by one cell stage with respect to transcriptional activation. The secreted ligand/inhibitor proteins were considered to be unstable and to signal for less than 30 minutes after production.

9.3 Principle of the cell contact areas model

Fig. 4 presents a global overview of the model, which aims at testing whether the areas of physical contact between cells emitting an extracellular ligand and their responding neighbours have a sufficient encoding potential to explain known fate restriction events during early ascidian embryogenesis.

To test this hypothesis, we made the following assumptions.

1. **Cell fate restriction events in ascidians are driven by cell communication and occur mainly through:** i) the polarization of a mother cell by extracellular signals emitted by its neighbours, followed by the differential inheritance of signaling cues by the two daughter cells (e.g. (32)); or ii) the reception of qualitatively or quantitatively different extracellular signals by the two daughters cells (e.g. (12)). Fate restriction thus results from the differential induction of sister cells either directly or via the polarization of their mother.
2. **All cells are equally competent to respond to extracellular signals emitted by their physical neighbours.** This hypothesis is consistent with the maternal expression of genes encoding the surface receptors and intracellular signaling pathway components, and with the uniform distribution of these mRNAs in eggs (61). We also hypothesize that local receptor densities are equal at all cell-cell interfaces, consistent with *in vitro* induction of explanted cells (80). Extracellular ligand availability is thus considered to be the main driver of inductions.
3. **Extracellular signaling ligands act at the interface of cells that secrete them (i.e. their lateral diffusion is negligible).** A signaling ligand can thus only activate receptors located at the physical interface with a cell expressing the ligand. This situation is expected for ligands associated with the plasma membrane (Delta, Ephrins). It was also proposed for FGF signaling in ascidian neural (12), mesenchyme (32) and notochord (81) inductions. Experiments in vertebrates or flies also suggested that Nodal, Bmp and Wnt ligands may, in certain biological contexts at least, only act at short range without significant lateral diffusion (82, 83). Ligands are considered to act in a juxtacrine (Ephrin, Delta) or autocrine/paracrine (FGF, BMP, WNT, Nodal) manner.
4. **Sequestering signaling inhibitors (Chordin, Noggin, ...) do not diffuse laterally and their action is dominant over signaling ligand:** if both a sequestering inhibitor and its target signaling ligand are present at a given cell interface, we consider that the inhibitor fully blocks the function

of the ligand at this interface. This is consistent with the finding that the interaction between Bmp4 and its inhibitor Chordin has a $K_D = 3.10^{-10}$ and equimolar amounts are sufficient to suppress BMP action (84).

5. **Ligand concentrations at cell-cell interfaces, receptor densities and ligand-receptor association constants are considered equal for all signaling pathways.** Consistently, *in vitro* induction assays suggest that nanomolar range concentrations of FGF, Bmp or Nodal are sufficient for cell inductions (FGF: (43, 85); Bmp: (77); Nodal: (86)). The *in vitro* association constants between Nodal, BMPs, FGFs, or Ephrins and their receptors are all in the range of $k_f = 10^5 \text{ s}^{-1} \cdot \text{mol}^{-1}$ (87–90).

With these simplifying assumptions, we define a signaling interface between two cells for a given pathway as an interface in which the ligand, but no inhibitor, is found. We then consider that the quantitative information deciding whether a cell is induced by a given pathway is provided by the sum of the areas of all signaling interfaces of the cell for this pathway, as the total number of receptors for this pathway on the receiving cell is proportional to the surface of contact between cells.

With these hypotheses, the model uses mass action kinetic equations to compute in each cell the concentration of activated intracellular effector for each pathway. It takes as inputs the binarized expression patterns of signaling extracellular ligands/antagonists, and the measured areas of contacts between cells in ASTEC-Pm1. The free parameters of the model are then explored to identify parameter combinations providing the best explanation of a set of 18 experimentally characterized ascidian embryonic inductions and 57 uninduced cells.

9.4 Formalization of the signaling models

Let us consider the case of a cell a , expressing a transmembrane receptor R, and an intracellular effector E. Binding of extracellular ligand L to the receptor R leads to the formation at the surface of a of an active ligand/receptor complex R^C , which in turns transforms the effector E into its activated form E^* at the inner side of the membrane of cell a . Depending on the pathway considered, different forms of interaction are considered (Fig. S33).

- **Simple induction.** In the simplest case, the presence of ligand L in the extracellular space is both necessary and sufficient to activate the transmembrane receptor R. This case is restricted in our model to the Notch pathway as no antagonist of this ligand is expressed at the stages considered (Fig. S33A).
- **Repression of the action of L by a secreted antagonist I,** emitted by cell c and preventing the interaction of the ligand with its receptor at the interface between cells a and c . This case is observed in the Bmp, Wnt and Nodal signaling pathways (Fig. S33B).
- **Repression of the action of L by the activation of a parallel pathway (L', R')** which activates an effector E' , in turn blocking the activation of a downstream effector E of the L pathway. This case is observed for the FGF pathway, whose intracellular activation of Ras is counteracted by the Ephrin pathway (Fig. S33C).

Within each class above, the model distinguishes two types of inductions.

- **Juxtacrine induction.** The ligand L displayed by cell b can only activate receptors harbored by a cell a in direct physical contact with b . Receptors on the surface of b cannot respond to ligands expressed by b . In our case, juxtacrine inductions are limited to transmembrane ligands of the Notch and Ephrin pathways.
- **Autocrine/paracrine induction.** The ligand L secreted by cell b can activate receptors located on both cell b and any cell a in direct contact with cell b , as we consider that lateral diffusion limits paracrine effects to direct physical contacts. All ligands of the FGF, Wnt, Nodal and Bmp pathways are considered to act in an autocrine/paracrine manner.

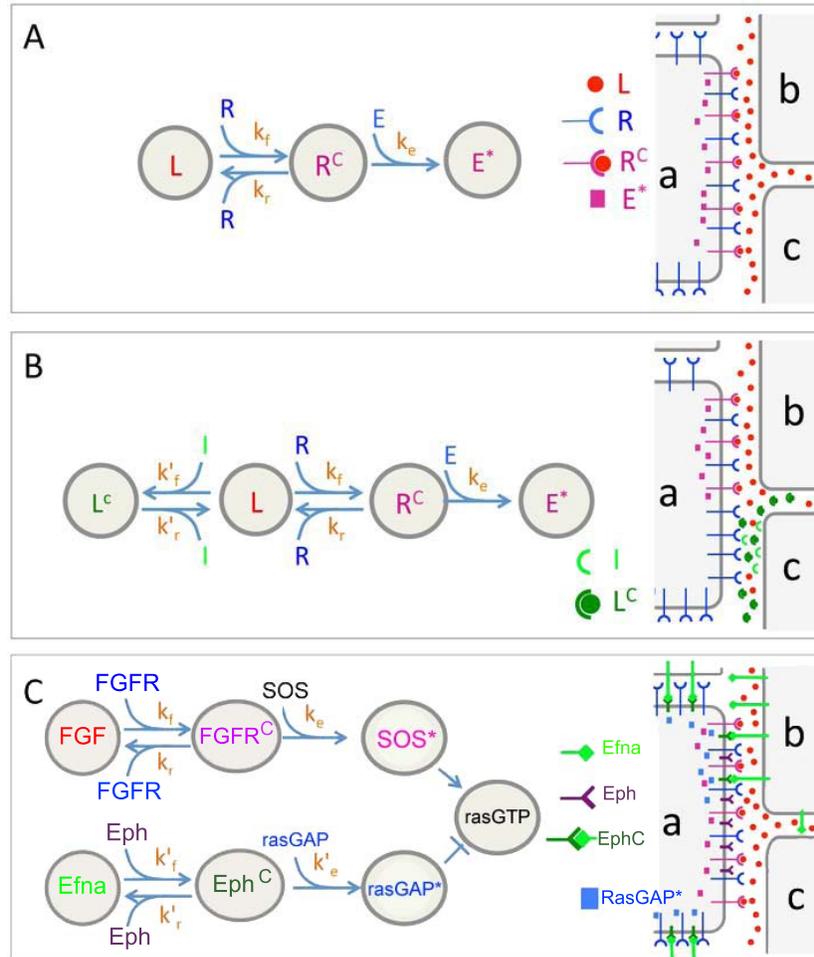


Figure S33: Types of signaling molecules involved and effect on inductions. A) simple induction. B) Repression of the action of L by a sequestering inhibitor I . In the actual model, we consider that sequestering inhibitor is dominant over the inducer and fully complexes it (L^c) at interfaces where both ligand and inhibitor are present. C) Repression of the action of FGF by the activation of a parallel pathway (Efna, Eph). In A and B, the ligands and inhibitors act very locally on the state of activation of their effectors. In C, RasGAP can diffuse away from its site of activation.

9.4.1 Computation of the positive signaling impact of a signaling ligand on an individual cell

Let us consider the case where a cell i responds to a signaling ligand L secreted by its neighbouring cells. The aim of this section is to compute the amount of intracellular effector E^* activated in cell i by

the ligand L secreted by the neighbouring cells of i .

Binding of L to R at the interface between the two cells leads to the formation of an active receptor-ligand complex R^C (association constant k_f , dissociation constant k_r), which catalyzes the activation of the effectors E to produce E^* (association constant k_e). The corresponding reactions are:



Let us now consider a cell i and its total membrane surface $A(t)$ exposed at time t to a given ligand L secreted by its neighbours (paracrine signaling) or by itself (autocrine signaling). Let $L(t)$ be the concentration of the ligand L in the intercellular space of volume v_{int} between cell i and all its neighbours emitting the ligand L [$\#$ molecules/ μm^3] and $R(t)$ and $R^C(t)$ be respectively the number of free and complexed receptors at the surface of cell i exposed to L [$\#$ molecules]. $E(t)$ and $E^*(t)$ are the number of inactive and active effectors in cell i at time t . The mass action kinetic equations for these reactions at the interface of the two cells are:

$$\frac{dR^C(t)}{dt} = k_f \cdot R(t) \cdot L(t) - k_r \cdot R^C(t), \quad (9.5)$$

$$\frac{dE^*(t)}{dt} = k_e \cdot R^C(t) \cdot E(t) - k_d R^C(t) E^*(t), \quad (9.6)$$

where $k_f [M^{-1}s^{-1}]$, $k_r [s^{-1}]$, $k_e [\#^{-1}s^{-1}]$, k_d and $[\#^{-1}s^{-1}]$ are constants. As we neglect here the lateral diffusion of molecules (Hypothesis 3 above) the conservation of mass imposes that the total (complexed and free) ligand and receptor concentrations or numbers respectively at the surface exposed to the ligand L, $L^T(t)$ and $R^T(t)$, satisfy:

$$v_{int} L^T(t) = v_{int} L(t) + R^C(t), \quad (9.7)$$

$$R^T(t) = R(t) + R^C(t), \quad (9.8)$$

$$E^T(t) = E(t) + E^*(t). \quad (9.9)$$

9.4.2 Number of receptors complexed at a cell membrane

As only a minority of receptors are activated (91), we first assume that $R^C(t)$ is much lower than $R(t)$. For the pathways considered $L(t)$ is in the nanomolar range (43, 77, 85, 86), in the order of, or larger than, $K_D = \frac{k_r}{k_f}$, which ranges from $0.6 \cdot 10^{-8}$ to 10^{-12} (87, 89, 90, 92, 93). Therefore $k_r \cdot R^C(t) \ll k_f \cdot R(t) \cdot L(t)$ and equation (9.5) can be simplified as:

$$\frac{dR^C(t)}{dt} = k_f \cdot R(t) \cdot L(t), \quad (9.10)$$

or, taking the conservation of mass into account:

$$\frac{dR^C(t)}{dt} = k_f \cdot [R^T(t) - R^C(t)] \cdot [L^T(t) - \frac{R^C(t)}{v_{int}}]. \quad (9.11)$$

Considering that $v_{int}L^T(t) \gg R^C(t)$ (most of the ligands in the intercellular space are not complexed with a receptor):

$$\frac{dR^C(t)}{dt} = k_f \cdot L^T(t) \cdot [R^T(t) - R^C(t)]. \quad (9.12)$$

We assume the local receptor density, α_R (receptors. μm^{-2}), at the surface of a cell is independent of the pathway, is the same for all cells and uniform on the surface of each cell (Hypotheses 2 and 5). The total number of receptors that could engage in signaling at the surface of cell i exposed to L, $R^T(t)$, is therefore defined by:

$$R^T(t) = \alpha_R \cdot A(t). \quad (9.13)$$

where $A(t)$ is the surface of contact (μm^2) with cells expressing the ligand but no inhibitor.

Replaced in equation (9.12), this leads to:

$$\frac{dR^C(t)}{dt} = k_f \cdot L^T(t) \cdot [\alpha_R \cdot A(t) - R^C(t)]. \quad (9.14)$$

We then consider that because of the short induction time in ascidians (typically a few minutes (31)), the total ligand concentrations are constant during the induction:

$$L^T(t) = L^T, \quad (9.15)$$

and as most cell contacts remain stable during the life of a cell (Fig. S27B), we approximated the area of contact with the mean value A of the total area of cell i exposed to L over the period of cell communication. Assuming the initial number of complexed receptors is null in every cell, $R^C(0) = 0$, equation (9.14) can be integrated, leading to:

$$R^C(t) = A \cdot \alpha_R \cdot [1 - e^{-k_f \cdot L^T \cdot t}]. \quad (9.16)$$

9.4.3 Number of effectors activated at a cell membrane

Using the mass conservation equation of effectors, and neglecting the degradation of the activated effectors, Eq. (9.6) becomes:

$$\frac{dE^*(t)}{dt} = k_e \cdot R^C(t) \cdot (E^T - E^*(t)) - k_d R^C(t) E^*(t). \quad (9.17)$$

Eq. (9.17) is of the form $\frac{dE^*}{dt} + P(t)E^* - Q(t) = 0$. Provided both $P(t)$ and $Q(t)$ are integrable functions, its general solution is:

$$E^*(t) = \frac{1}{e^{\int_0^t P(\tau) d\tau}} \left[\int_0^t e^{\int_0^s P(\tau) d\tau} Q(s) ds \right], \quad (9.18)$$

where we have imposed the boundary condition $E^*(0) = 0$. As $Q(t) = \frac{k_e E^T}{k_e + k_d} P(t)$, Eq. (9.18) reduces to

$$\begin{aligned} E^*(t) &= \frac{k_e E^T}{k_e + k_d} e^{-\int_0^t P(\tau) d\tau} \left[\int_0^t e^{\int_0^s P(\tau) d\tau} P(s) ds \right] = \frac{k_e E^T}{k_e + k_d} e^{-\int_0^t P(\tau) d\tau} \left[\int_0^t \frac{d}{ds} \left(e^{\int_0^s P(\tau) d\tau} \right) ds \right] \\ &= \frac{k_e E^T}{k_e + k_d} e^{-\int_0^t P(\tau) d\tau} \left(e^{\int_0^t P(\tau) d\tau} - 1 \right) = \frac{k_e E^T}{k_e + k_d} \left(1 - e^{-\int_0^t P(\tau) d\tau} \right). \end{aligned} \quad (9.19)$$

Since $P(t) = k_e R^C(t)$ one finally obtains

$$E^*(t) = \frac{k_e E^T}{k_e + k_d} \left(1 - e^{-k_e \int_0^t R^C(\tau) d\tau} \right). \quad (9.20)$$

$R^C(t)$ being a chemical concentration and, as such, non-negative, the integral at the exponent in the r.h.s. of Eq. (9.20) is itself non-negative for $\forall t > 0$ and $E^*(t)$ is upper-bounded by $\frac{k_e E^T}{k_e + k_d}$, which also represents its asymptotic limit since $\lim_{t \rightarrow \infty} \int_0^t R^C(\tau) d\tau = \infty$. Eq. (9.20) is fully general, but its explicit form depends on $R^C(t)$ explicited in Eq. (9.16). One therefore finally obtains:

$$E^*(t) = \frac{k_e E^T}{k_e + k_d} \left(1 - e^{-AF(t)} \right), \quad (9.21)$$

where

$$F(t) = \frac{\alpha_R k_e}{k_f L^T} \left[k_f L^T t - \left(1 - e^{-k_f L^T t} \right) \right]$$

is a positive function of t that does not depend on A . $E^*(t)$ thus monotonically increases with time for all possible values of $k_f \cdot L^T$ or $\alpha_R \cdot k_e$.

Finally, k_e , k_d and E^T being constants, we consider the normalized activated effector value \bar{E}^* :

$$\bar{E}^*(t) = \frac{E^*(t) k_e + k_d}{E^T k_e} = \left(1 - e^{-AF(t)} \right) \quad (9.22)$$

Interestingly, \bar{E}^* is a non-linear, monotonically-increasing function of the surface A of cell i exposed to ligand L, which saturates to 1 with increasing A . The number of effectors activated by a ligand L thus increases (quasi-linearly within the range of biologically relevant surfaces) with the area of contact to ligand expressing cells (Fig. S34).

9.5 Computation of the combined signaling impacts of a ligand and its secreted antagonist on an individual cell

We now consider the case of a cell exposed to both a ligand L and its secreted antagonist I, as observed for the BMP, Nodal and Wnt pathways. The antagonist is considered to act by preventing the binding of the ligand to its receptor. Following hypothesis 4 of the model, we consider that if both L and I are present at the interface between two cells, the inhibitor prevents all available ligands from interacting with their cognate receptors at this interface (Fig. S33B).

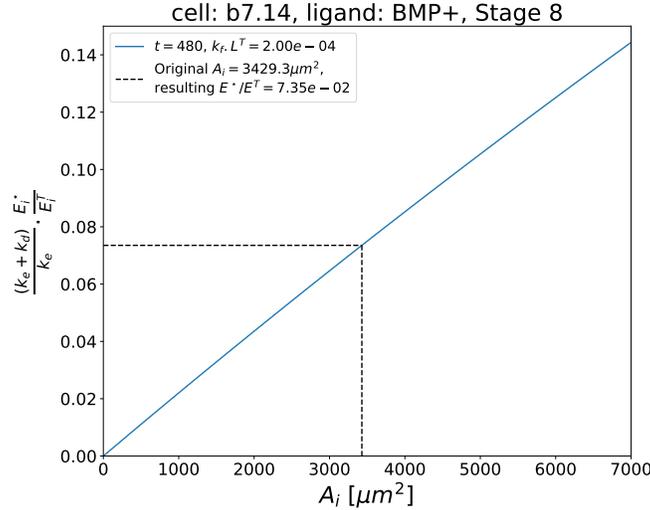


Figure S34: Example of the relationship between $\frac{(k_e+k_d)}{k_e} \cdot \frac{E_i^*}{E_i^T}$ and the area of contact A_i between emitting and responding cells. This graph is computed for the E_i^* of BMP in the cell b7.14 with $k_f \cdot L^T = 2e^{-4} \text{ s}^{-1}$ and induction time of 480 s.

The action of the inhibitor on a cell i can thus be modelled by considering that the surface of contact established by cell i with cells j expressing ligand L (including i itself if applicable) only contributes to the induction if neither cell i nor cell j express I. Let the set $N_i(L)$ be the set of neighbouring cells that are considered for the expression of ligand L to the cell i . Then we can compute A as following:

$$A = \sum_{j \in N_i(L)} A_{ij}, \quad (9.23)$$

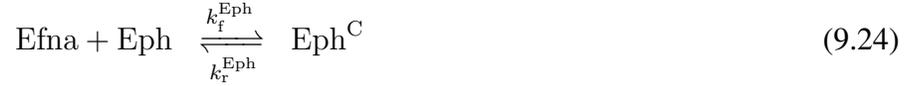
where A_{ij} is the surface between the cells i and j exposed to L but not I.

In our implementation, A_{ij} is computed differently according to whether we are considering an induction or a polarization. In the induction case, the area A_{dj} between a daughter cell d and cell j is its average from 4 minutes to 4 + t minutes after the division, t being the time of induction. In the case of polarization, the surface of contact A_{Mj} between a mother cell M and a cell j that will belong, after the division, to one of the daughter cells d , is the average of the area of contact between d and j , $A_{dj}(t)$, from 4 to 14 minutes after the division of M .

9.6 The case of mutually antagonistic FGF/Ephrin signaling

The FGF and Ephrin pathways both control, in opposite ways, the activation of the Ras protein at the top of the ERK signaling pathway. Binding of FGF to its receptor leads to the activation by recruitment to the membrane of the SOS Ras-GEF (94). Binding of EphrinA to its Eph receptor activates p120RasGAP (95) (Fig. S33C). As the integration of the two pathways occurs intracellularly, we considered that the Ephrin and FGF pathways do not need to be activated on the same cell interface, consistent with the work of (96) on neural induction. We thus consider the following reactions, adapted from (9.3) and (9.4):

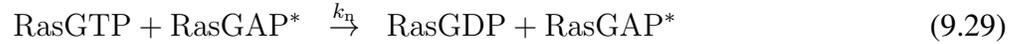
RasGap effector activation pathway:



SOS effector activation pathway:



In both pathways, the amount of effector is determined in Eq. (9.21) (note that the surfaces of contact with Ephrin- or FGF-expressing cells have no reason to be equal). The Efna ligand expressed by neighboring cells, acts in a juxtacrine way and forms a complex Eph^{C} with receptors Eph at the surface of cell i , thereby activating the RasGAP effector (association/dissociation constants k_a and k_b respectively). Similarly, the level of activation of SOS is defined in response to FGF signaling (association/dissociation constants k_i and k_j respectively). The effect of FGF is autocrine/paracrine. SOS^* and RasGAP^* regulate the balance between RasGDP and RasGTP at the inner plasma membrane:



In addition to p120-RasGAP, the ascidian genome contains another maternally expressed rasGAP molecule (KH.C1.339), which may also participate to the conversion of RasGTP into RasGDP, independently of Ephrin signaling. We formalized the existence of possible additional maternal GAP proteins, X^* , by adding an additional pathway transforming RasGTP into RasGDP:



leading altogether to the following RasGTP dynamics:

$$\frac{d\text{RasGTP}(t)}{dt} = k_m \text{RasGDP}(t) \text{SOS}^*(t) - \text{RasGTP}(t) (k_n \text{RasGAP}^*(t) + k_p X^*(t)). \quad (9.31)$$

These transitions are assumed to be very rapid compared with the extracellular processes leading to the activation of the receptors by their ligands. We therefore consider that a quasi-steady state is very rapidly reached for the final activated effector RasGTP:

$$\frac{d\text{RasGTP}(t)}{dt} = 0. \quad (9.32)$$

Assuming that the total amount of RasGTP/RasGDP is constant in the system:

$$\text{Ras}^{\text{T}} = \text{RasGDP}(t) + \text{RasGTP}(t), \quad (9.33)$$

we thus have:

$$0 = k_m \text{Ras}^T \text{SOS}^*(t) - \text{RasGTP}(t)(k_m \text{SOS}^*(t) + k_n \text{RasGAP}^*(t) + k_p X^*(t)), \quad (9.34)$$

leading to:

$$\text{RasGTP}(t) = k_m \cdot \text{Ras}^T \frac{\text{SOS}^*(t)}{(k_m \text{SOS}^*(t) + k_n \text{RasGAP}^*(t) + k_p X^*)}, \quad (9.35)$$

and to the following expression of $\text{RasGTP}(t)$ in the quasi-steady state:

$$\text{RasGTP}(t) = \text{Ras}^T \frac{\text{SOS}^*(t)}{(\text{SOS}^*(t) + \delta_1 \text{RasGAP}^*(t) + \delta_2)}, \quad (9.36)$$

where $\delta_1 = \frac{k_n}{k_m}$ and $\delta_2 = \frac{k_p X^*}{k_m}$ are constants (the amount of maternal effectors X^* is considered to be constant).

Finally, as Ras^T is a constant across stages and cells, the expression of the normalized fraction of $\text{RasGTP}(t)$, $\overline{\text{RasGTP}}(t)$ can be expressed as:

$$\overline{\text{RasGTP}}(t) = \frac{\text{RasGTP}(t)}{\text{Ras}^T} = \frac{\text{SOS}^*(t)}{(\text{SOS}^*(t) + \delta_1 \text{RasGAP}^*(t) + \delta_2)} \quad (9.37)$$

9.7 Deciding whether an induction takes place

9.7.1 Definitions

For each value of the five free parameters $(t, k_f \cdot L^T, \alpha_R \cdot k_e, \delta_1, \delta_2)$, the preceding equations allow the computation of the normalised amount of activated effector $\overline{E}^*(t)$ in each cell and for each pathway. Clearly, the value of \overline{E}^* is a function of both time and of the specific cell i analyzed, such that $\overline{E}^* = \overline{E}^*(t, i)$. Based on these values, we describe in this section, the method we designed to decide whether sister cells are differentially induced by a specific signaling pathway. In what follows, with a slight abuse of notation, we generically refer to these normalised molecular amounts \overline{E}^* as *amount of activated effectors*.

We introduce now an additional free parameter in the differential induction model, namely, the time t_0 during which a cell is exposed to one or several inducers. We assume this parameter to be common to all induction pathways. The relevant quantity to predict whether a cell i is induced is therefore the value $\overline{E}^*(t_0, i)$. Hereafter, t_0 being fixed and equal for all cells and pathways, we will use the simplifying notation $\overline{E}^*(t_0, i) = \overline{E}^*(i)$.

Let $\overline{E}^*(s_1)$ and $\overline{E}^*(s_2)$ be the amounts of an activated effector respectively estimated by the model in two sister cells s_1 and s_2 for a considered pathway, and such that $\overline{E}^*(s_1) > \overline{E}^*(s_2)$.

Induction of cells. We decide that a cell is induced if, after a time t_0 of exposure to the inducers of a pathway, its amount of activated intracellular effectors for this pathway is higher than a threshold ϵ_{max} . Symmetrically, we consider that, if the amount of activated intracellular effectors is lower than a non-induction threshold ϵ_{min} at time t_0 , the cell is uninduced. For consistency we impose:

$$\epsilon_{min} < \epsilon_{max}. \quad (9.38)$$

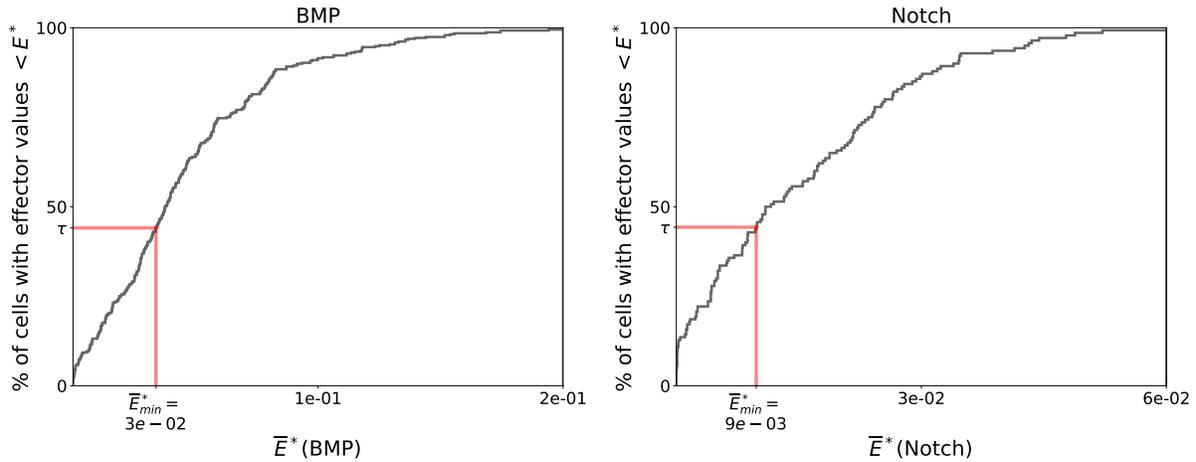
If at t_0 , the amount of activated intracellular effectors lies between ϵ_{min} and ϵ_{max} , we consider that we cannot conclude on the induction status of the cell.

Differential induction of sister cells. We consider that two sister cells s_1 and s_2 are differentially induced by a given pathway if s_1 is induced while s_2 is not, that is, the normalized activated effector amounts $\bar{E}^*(s_1)$ and $\bar{E}^*(s_2)$ should be respectively above ϵ_{max} and below ϵ_{min} for this pathway. If the induction of one of the two cells remains undetermined, we decide to apply a conservative rule by considering that the two sister cells are not differentially induced.

9.7.2 Algorithm to estimate of ϵ_{max} and ϵ_{min} for each pathway

To determine plausible values of these parameters, we first estimate an approximate value, \bar{E}_{min}^* , below which a cell is sure not to be induced, which we refine by assessing differential induction candidates and by using consistency rules to compute ϵ_{max} and ϵ_{min} for each pathway.

Minimal signal intensity \bar{E}_{min}^* for a cell to be induced. Differential induction of a sister pair by a given pathway requires that one of the sisters receives a sufficient signaling intensity to be induced. We first define \bar{E}_{min}^* as the minimal effector amount for a cell to possibly be induced by this pathway. The cells that receive the weakest signals from ligands in this pathway are unlikely to be induced. Conversely, the cells that receive the strongest signal are very likely to be induced. To extend this reasoning, we consider that a cell can only be induced in a given pathway if it is among the τ % of cells receiving the strongest signal from this pathway. This strategy allows to define a single τ free parameter value for all pathways, which translates into different threshold values \bar{E}_{min}^* for each pathway and stage (Fig. S35).



(a) Cumulative distribution of BMP signaling received by cells (b) Cumulative distribution of Notch signaling received by cells

Figure S35: Defining a minimal level of signaling necessary for inductions to occur. Note that a same τ value translates into different \bar{E}_{min}^* values for the different pathways and stages, (the induction time used here was $t_0 = 10$ min.)

Initial set of differentially induced cells. Intuitively we expect that, in most cases, two sister cells will be differentially induced if they receive markedly different induction cues. We therefore consider, as a first approximation, that $\bar{E}^*(s_1)$ and $\bar{E}^*(s_2)$ should be markedly different for this pathway. To formalize this condition we introduce the variable

$$\rho = \bar{E}^*(s_1)/\bar{E}^*(s_2). \quad (9.39)$$

All the sister pairs in which at least one sister has an activated effector amount greater than E_{min}^* and whose effector ratio ρ is greater than a certain threshold ρ_0 are considered candidates for differential induction (Fig. S36). ρ_0 is assumed to be common to all pathways and developmental stages and represents a free parameter of the model. Let us call \mathcal{S}_{ρ_0} this set of candidate pairs.

Determination of ϵ_{min} and ϵ_{max} . We then refine this first selection \mathcal{S}_{ρ_0} by checking induction consistency. Because all cells are considered equally competent to respond to a given signal (Hypothesis 2 of the model), the amount of activated effectors in any induced cell should be higher than the amount of activated effectors in any uninduced cell. To be consistent, we therefore have to impose that all cells with an activated effector amount lower than the highest $\bar{E}^*(s_2)$ for an uninduced cells, have to be uninduced, and therefore that the threshold value under which cells are sure not to be induced is:

$$\epsilon_{min} = \max_{(s_1, s_2) \in \mathcal{S}_{\rho_0}} \bar{E}^*(s_2). \quad (9.40)$$

We should therefore eliminate from \mathcal{S}_{ρ_0} all pairs where the candidate induced sister (s_1) has an activated effector amount lower than ϵ_{min} . This defines the set \mathcal{S}'_{ρ_0} :

$$\mathcal{S}'_{\rho_0} = \{(s_1, s_2) \in \mathcal{S}_{\rho_0} | \bar{E}^*(s_1) > \epsilon_{min}\}. \quad (9.41)$$

Then, the threshold above which induction arises certainly is defined by:

$$\epsilon_{max} = \min_{(s_1, s_2) \in \mathcal{S}'_{\rho_0}} \bar{E}^*(s_1). \quad (9.42)$$

The consistency condition is reflected by:

$$\epsilon_{max} \geq \epsilon_{min}. \quad (9.43)$$

9.7.3 Deciding on differential induction of sister cells.

Based on these values of ϵ_{min} and ϵ_{max} , cells can be classified into three sets:

- Uninduced $\mathcal{U} = \{s \in C, \bar{E}^*(s) \leq \epsilon_{min}\}$
- Induced $\mathcal{I} = \{s \in C, \epsilon_{max} \leq \bar{E}^*(s)\}$
- Undetermined $\mathcal{G} = \{s \in C, \epsilon_{min} < \bar{E}^*(s) < \epsilon_{max}\}$.

A couple of sister cells (s_1, s_2), $E^*(s_2) < E^*(s_1)$ is therefore considered **differentially induced** if and only if $s_1 \in \mathcal{I}$ and $s_2 \in \mathcal{U}$. Note that cells satisfying this condition may not necessarily satisfy equation (9.39), which only constituted a first order approximation.

9.8 Training the model

9.8.1 A census of known ascidian inductions

We scanned the literature to identify cases of induction where the type of action and the identity of the inducer was known. *Phallusia mammillata* being, like the major ascidian model *Ciona intestinalis*, a Phlebobranch, we considered that induction data from *Ciona intestinalis* could be extrapolated to *Phallusia* (Table S11).

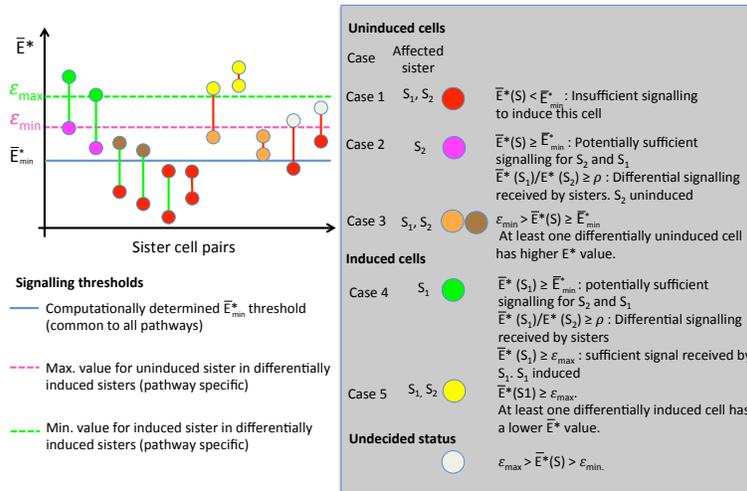


Figure S36: Signaling thresholds.

9.8.2 Exploration of parameter space

We next ran the model from developmental stage 8 to 12. In this developmental stage range, the exhaustive atlas of signaling gene expression was available.

To identify optimal sets of parameters, we screened the parameter space to find a combination that would correctly predict the following groups of cells:

- 18 known differential inductions (Table S11): This population includes 18 mother cells known from the literature to give rise to two differently fated daughters by either inductive polarization of the mother or differential induction of the sisters. For ten of these mother cells, the precise inducer and timing/mode of induction is well understood, identifying 14 combinations of inducing time/ligands. For the remaining eight inductions, the inducer is known, but not its precise timing of action.
- 49 cell divisions not giving rise to differentially induced cell pairs (a6.5, a6.6, A7.1, A7.2, A7.3, A7.5, A7.7, a7.11, a7.12, a7.14, a7.15, a7.16, A8.1, A8.2, A8.3, A8.4, A8.5, A8.6, A8.9, A8.11, A8.13, A8.14, a8.21, a8.22, a8.23, a8.24, a8.26, a8.27, a8.28, a8.29, a8.30, a8.31, a8.32, B7.1, B7.4, B7.5, B7.8, B8.4, B8.7, B8.8, B8.14, B8.15, B8.16, b8.20, b8.24, b8.26, b8.27, b8.29, b8.30). All the cells of this set had daughters that shared the same fate.
- And a third population that consisted of the remaining cells, whose induction status was unknown.

We screened the parameter space by dichotomous search and by scoring the quality of the model given a set of parameters. This score is the sum of twice the number of false positives (cells in the uninduced population for which differential inductions were detected), plus the number of missed inductions (cells in the induced population, for which no differential induction was detected), plus four times the number of missed cell/ligand combinations (experimentally determined inducing ligand not found in the prediction). The parameters then considered to fit best the ground truth are the ones which minimize this score.

The parameter space we explored was the following:

- ρ_0 varies from 1.6 to 2.6 by increments of 0.2 ((12) reported a 2 fold difference in surface of contact discriminated induced from uninduced cells during animal neural induction).
- τ from 38 to 50% by increments of 2 (meaning that a substantial minority of cells can be induced by a given signal).
- $k_f.L_T$ from 10^{-6} to $10^{-2} s^{-1}$ by geometric increments of 1.3 (for typical values of L_T and k_f , see section 9.3, Hypothesis 5).
- Time of induction from 4 to 14 minutes by increments of 2 minutes ((31) reported a 10 minutes exposure to FGF was sufficient for ascidian notochord induction).
- δ_1 from 1 to 10^3 by geometric increments of 3.16 (RasGAP transforms RasGTP into RasGDP faster than SOS transforms RasGDP into RasGTP (97)).
- δ_2 at different stages 6 and 8 together and 10 (therefore named $\delta_{6,8}$ and δ_{10} from now on) and both varying from 10^{-4} to 10^{-1} by geometric increments of 3.2 (this value is much smaller than the typical concentration of activated effector molecules in $\# \text{ mol} \cdot \mu\text{m}^{-3}$ in a cell).
- $\alpha_R.k_e$ from 10^{-8} to $10^{-4} \mu\text{m}^{-2} s^{-1}$ per cell by geometric increments of 2.8 (Receptor densities are typically $10 - 100 \# \cdot \mu\text{m}^{-2}$, k_e is in the order of $10^5 M^{-1} s^{-1}$ or $\approx 10^{-7} \#^{-1} s^{-1}$ per cell).

9.9 Description of the model results in wild-type conditions

Under wild-type conditions (meaning no geometrical, topological or genetic perturbation), the combination of parameters that minimizes the previously described score is shown in Table S12.

The results of the model run on the three different embryos with their optimal set of parameters are exposed in Fig. S37, S38, S39 and in Table S13. For example, when run on ASTEC-Pm1, the model misses none of the 14 ground truth differential cell induction events (defined by a triplet: cell pair/pathway/time), none of the 18 sister cells known to be differentially induced and finds potential inductions for 12 sister cell pairs not thought to be differentially induced (overpredicted differential inductions, false positives).

Fig. S40 to S57, present the predicted \bar{E}^* for each pathway and stage of analysis.

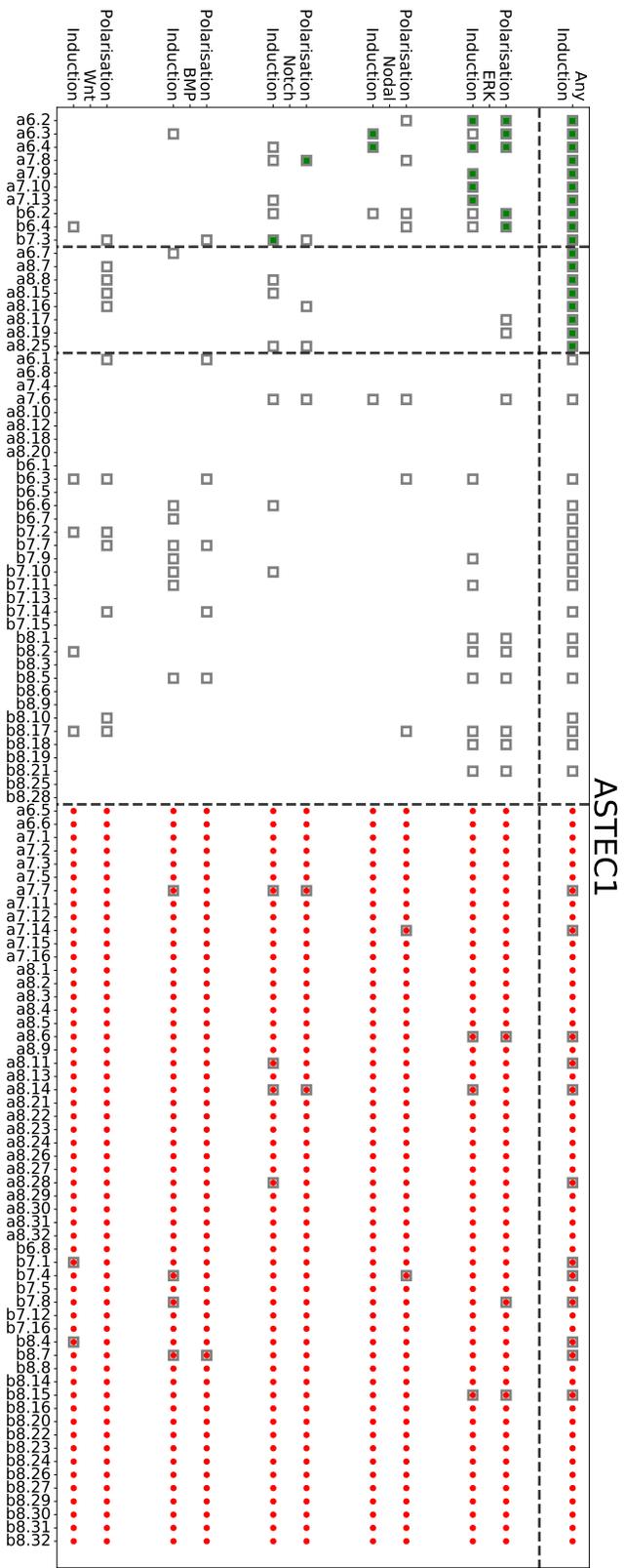


Figure S37: Result for the model run under wild type conditions for ASTEC-Pml1 This graphic shows, for each division event (named after the mother cell) the output of the model for every single pathway together with the ground truth. The green squares and red circles represent the ground truth. The green squares are the experimentally described differential inductions. The red squares indicate cells whose daughters are, not differentially induced. The grey squares represent for each pathway the mothers cells predicted by the model to be polarized or to produce differentially induced daughters.

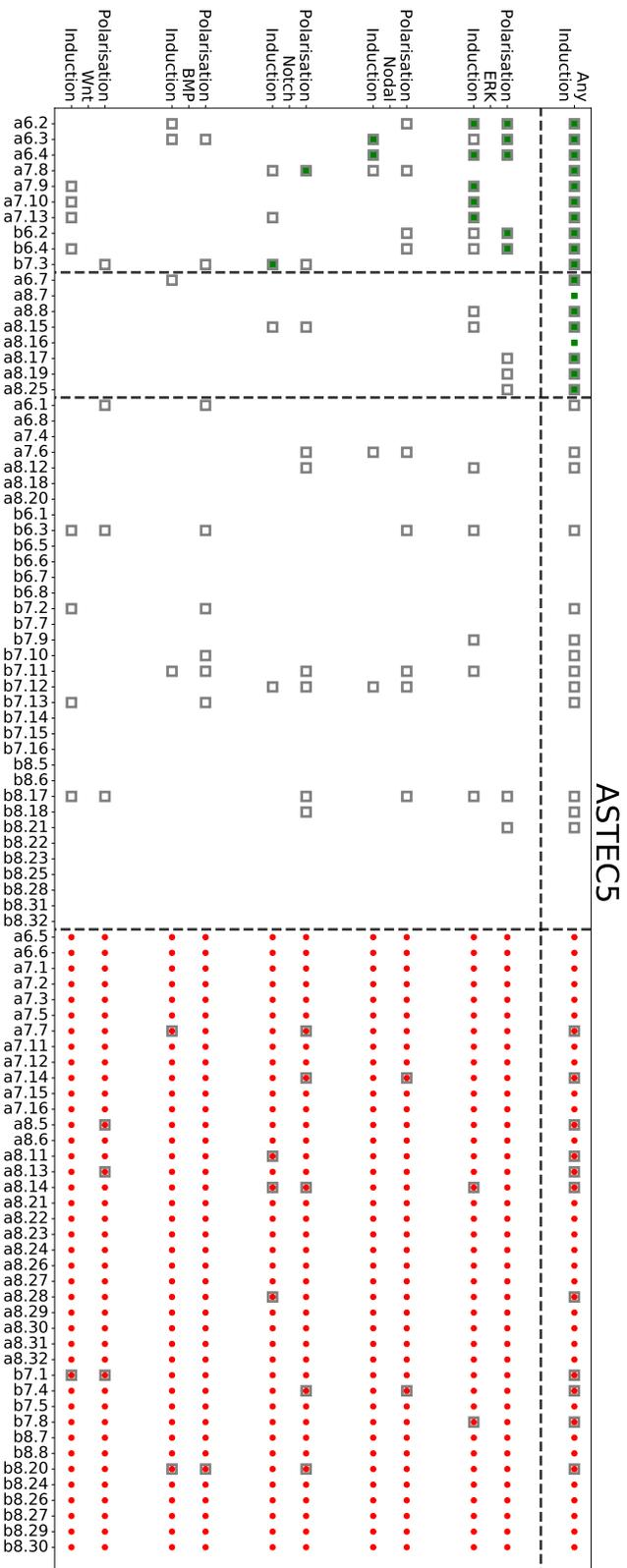


Figure S38: Result for the model run under wild type conditions for ASTEC-Pm5. This graphic shows, for each division event (named after the mother cell) the output of the model for every single pathway together with the ground truth. The green squares and red circles represent the ground truth. The green squares are the experimentally described differential inductions. The red squares indicate cells whose daughters are, not differentially induced. The grey squares represent for each pathway the mothers cells predicted by the model to be polarized or to produce differentially induced daughters. Note that compared to the two other embryos some cells are missing, this is due to the fact that this movie is shorter. The missing cells are cells that are not present in the recorded period of this embryo.

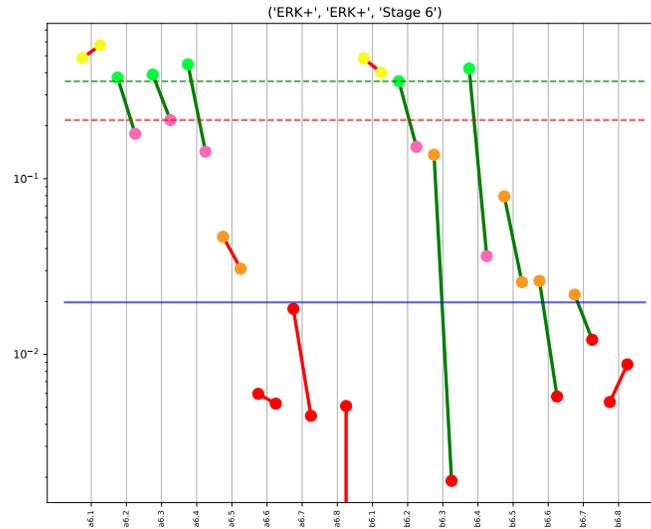


Figure S40: \bar{E}^* distribution, ERK, stage 6. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 6 to polarize the mother.

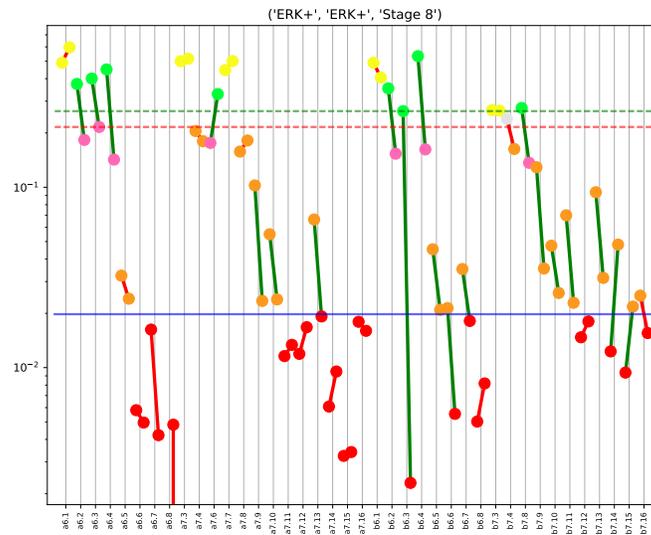


Figure S41: \bar{E}^* distribution, ERK, stage 8. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 7th generation mothers, or to induce the daughters of 6th generation mothers. ASTEC-Pm1 dataset.

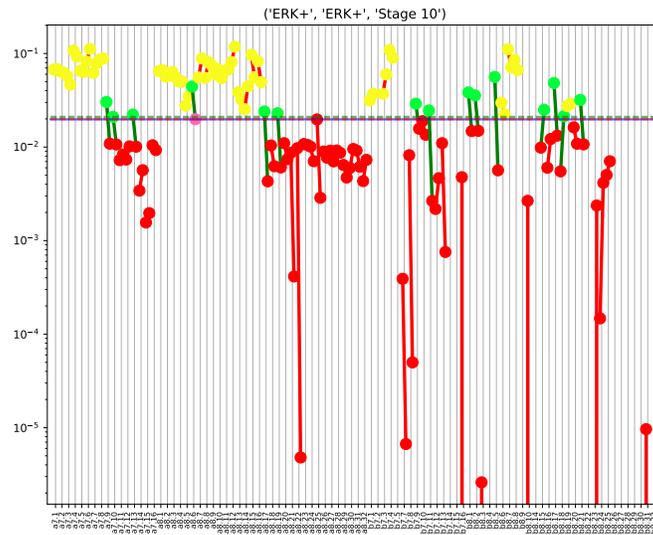


Figure S42: \bar{E}^* distribution, ERK, stage 10. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 10 and 11. ASTEC-Pm1 dataset.

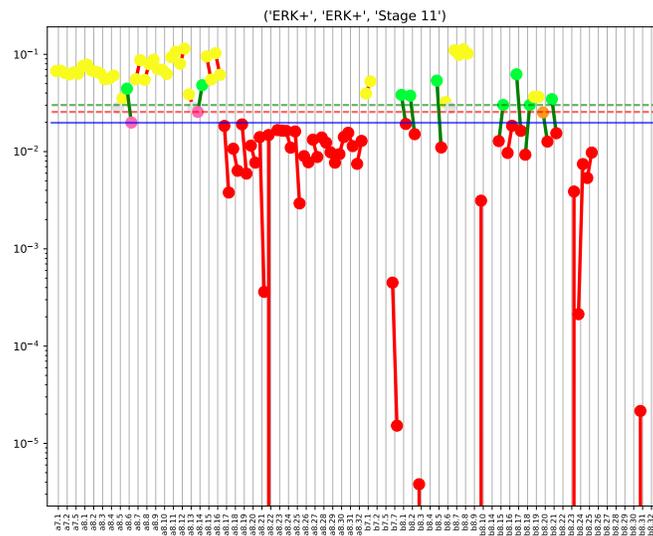


Figure S43: \bar{E}^* distribution, ERK, stage 11. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 11 and 12. ASTEC-Pm1 dataset.

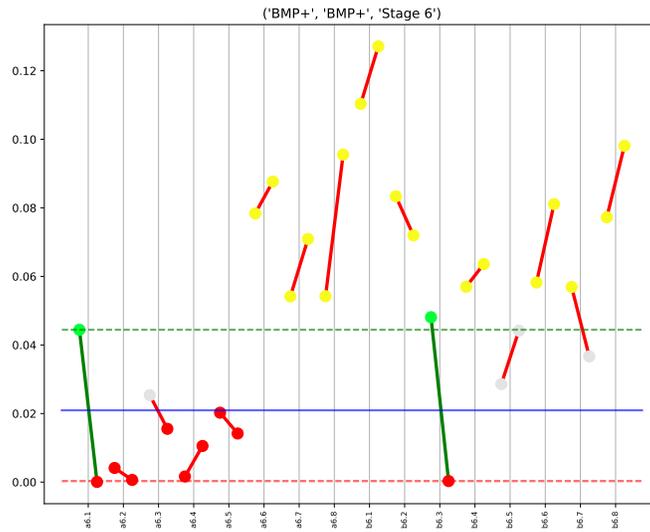


Figure S44: \bar{E}^* distribution, *Bmp*, stage 6. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 6 to polarize the mother. *ASTECPm1* dataset.

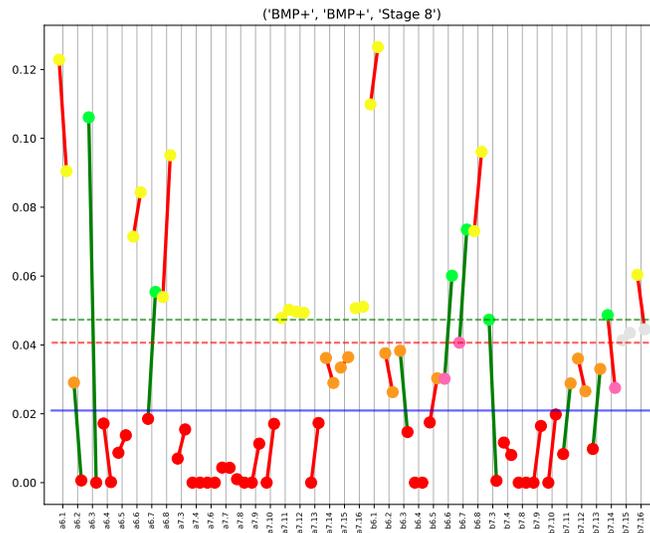


Figure S45: \bar{E}^* distribution, *Bmp*, stage 8. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 7th generation mothers, or to induce the daughters of 6th generation mothers. *ASTECPm1* dataset.

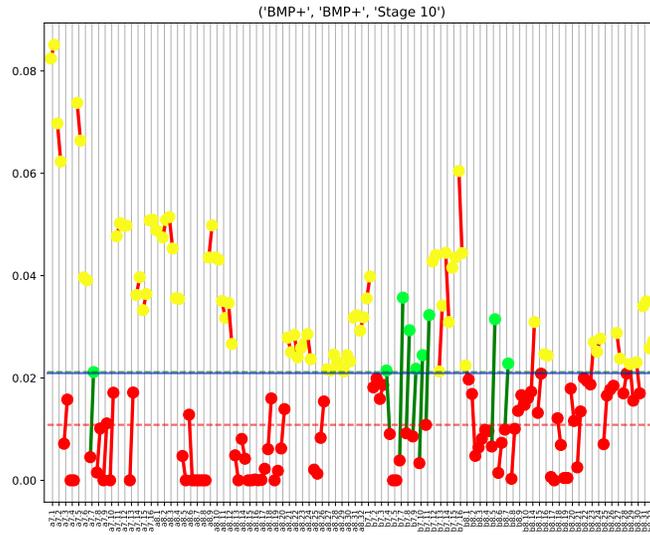


Figure S46: \bar{E}^* distribution, *Bmp*, stage 10. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 10 and 11. ASTEC-Pm1 dataset.

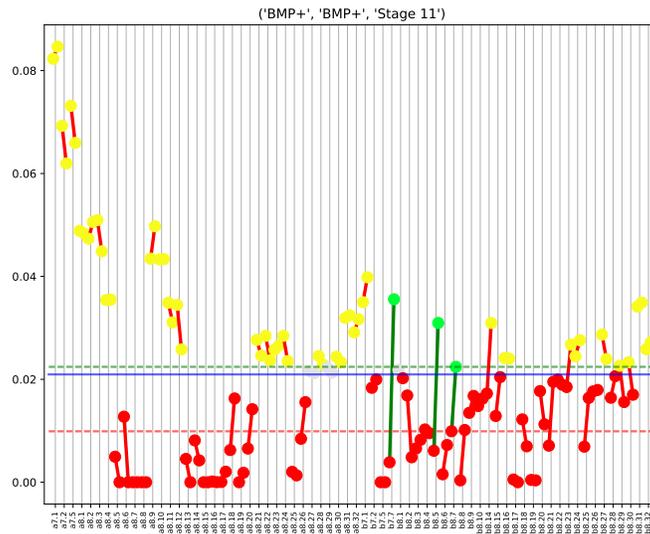


Figure S47: \bar{E}^* distribution, *Bmp*, stage 11. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 11 and 12. ASTEC-Pm1 dataset.

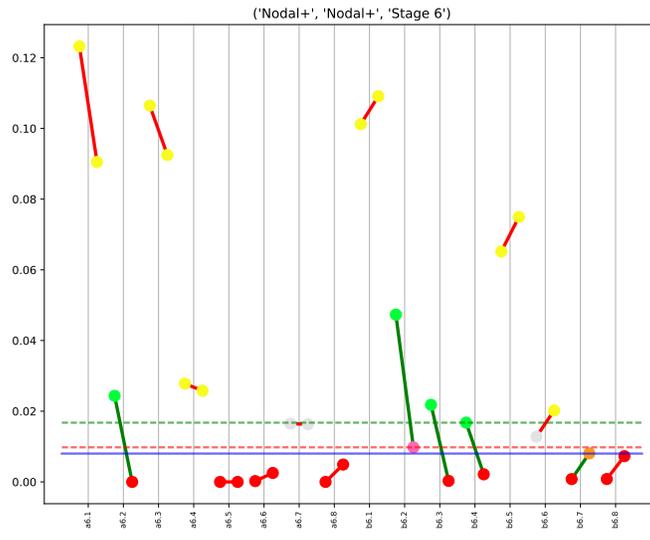


Figure S48: \bar{E}^* distribution, *Nodal*, stage 6. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 6 to polarize the mother. *ASTECPm1* dataset.

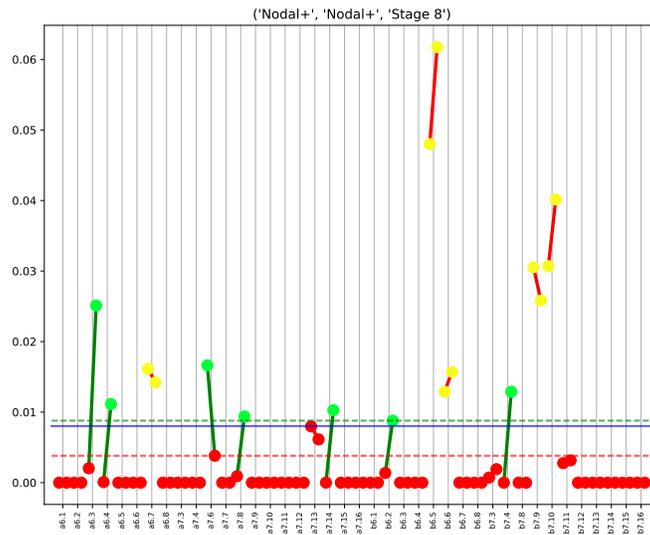


Figure S49: \bar{E}^* distribution, *Nodal*, stage 8. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 7th generation mothers, or to induce the daughters of 6th generation mothers. *ASTECPm1* dataset.

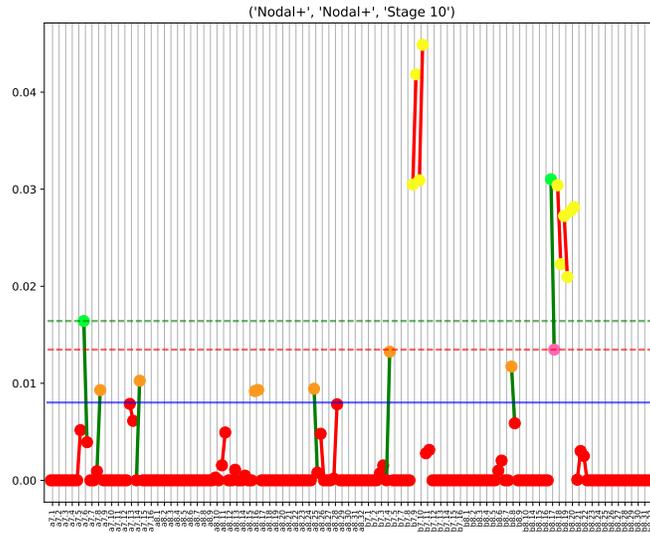


Figure S50: \bar{E}^* distribution, Nodal, stage 10. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signalling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 10 and 11. ASTEC-Pm1 dataset.

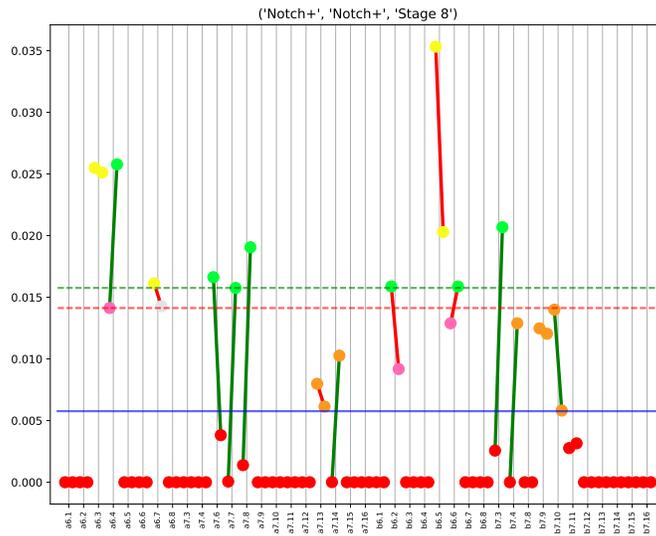


Figure S51: \bar{E}^* distribution, Notch, stage 8. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 7th generation mothers, or to induce the daughters of 6th generation mothers. ASTEC-Pm1 dataset.

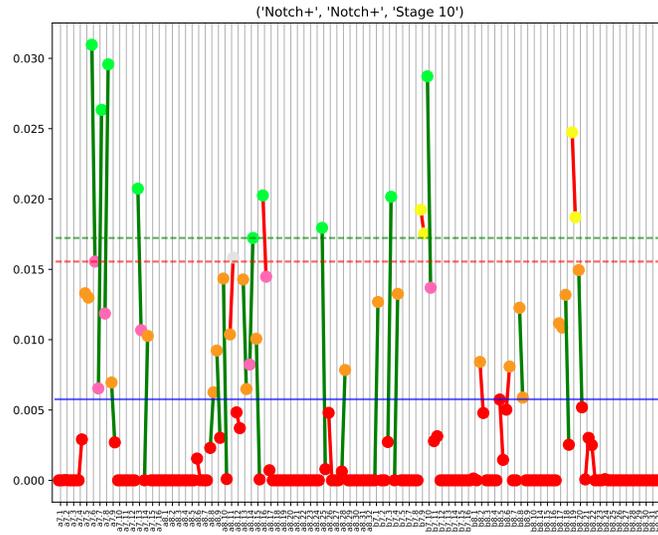


Figure S52: \bar{E}^* distribution, Notch, stage 10. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 10 and 11. ASTEC-Pm1 dataset.

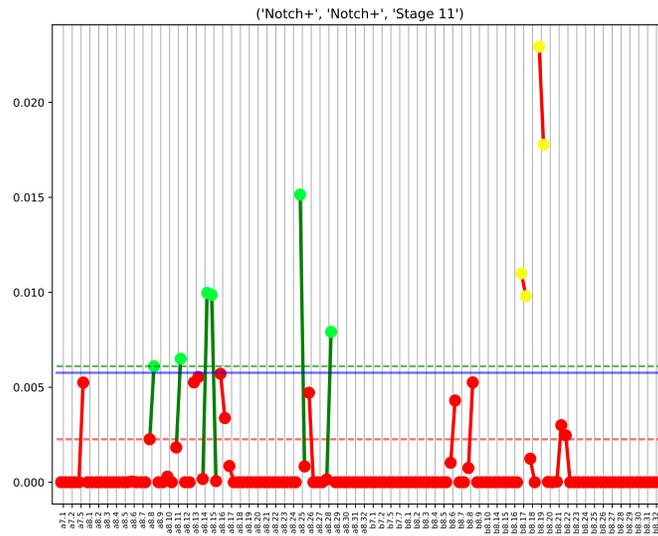


Figure S53: \bar{E}^* distribution, Notch, stage 11. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 11 and 12. ASTEC-Pm1 dataset.

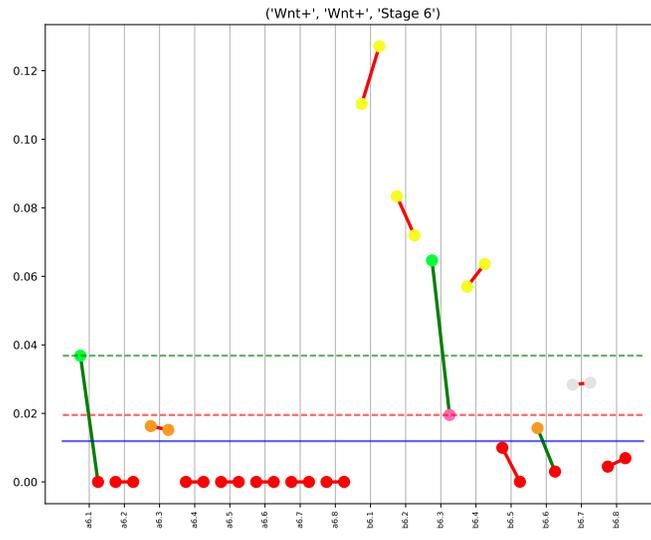
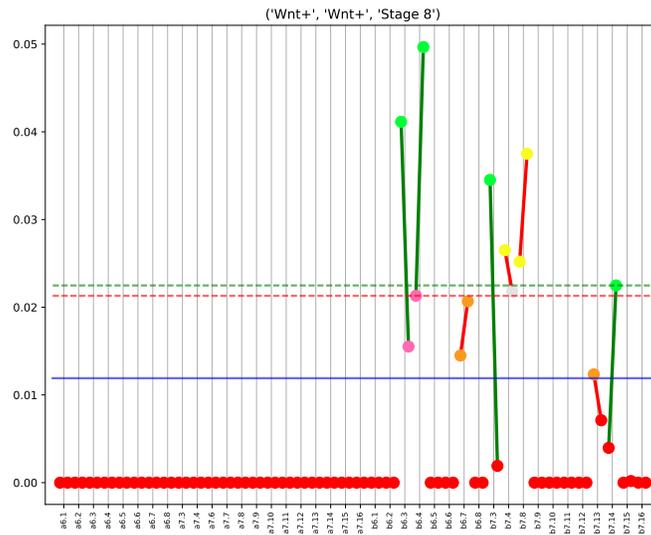


Figure S54: \bar{E}^* distribution, *Wnt*, stage 6. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 6 to polarize the mother. ASTEC-Pm1 dataset.



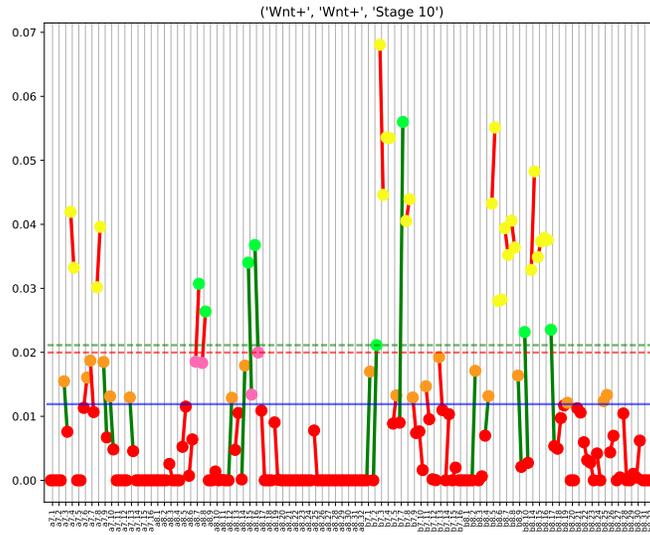


Figure S56: \bar{E}^* distribution, *Wnt*, stage 10. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 10 and 11. ASTEC-Pm1 dataset.

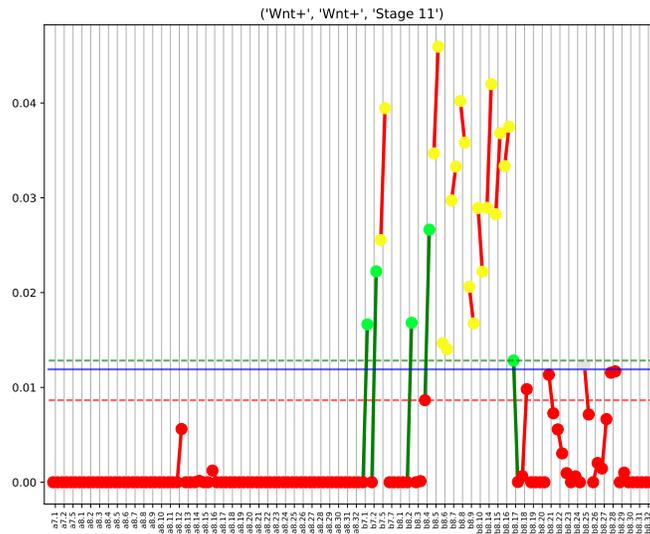


Figure S57: \bar{E}^* distribution, *Wnt*, stage 11. The graph indicates for each mother cell (x axis), the values of \bar{E}^* (Y axis) in each daughter cell (cell of smaller index to the left) as a result of signaling occurring at stage 8 to polarize 8th generation mothers, or to induce the daughters of 7th generation mothers cleaving between stage 11 and 12. ASTEC-Pm1 dataset.

9.10 Impact on the model output of perturbations in ligand expression profiles and embryo geometry

In section, we detail how we tested the sensitivity of the model to variations in different classes of input parameters.

9.10.1 Long range ligands and inhibitors action

We first tested the impact of increasing the range of action of secreted ligands and inhibitors (FGF, Bmp, Nodal and Wnt pathways) (see section 9.3), while keeping juxtacrine signaling by membrane-tethered ligands (Delta, Ephrin). The action range was set to second degree neighbours, meaning that a secreted molecule expressed by a given cell acts on both the direct physical neighbours of this cell as well as on the physical neighbours of these neighbours. We consider a long range action to second degree neighbours only if the surface of contact was sufficiently large and long lasting: let \mathcal{N}_c be the set of neighbors of the cell c , c_i be a cell expressing a ligand L , let c_j be a neighbor of c_i and c_k be a neighbor of c_j but not c_i :

$$c_j \in \mathcal{N}_{c_i} \quad (9.44)$$

$$c_k \in \mathcal{N}_{c_j}, c_k \notin \mathcal{N}_{c_i} \quad (9.45)$$

Then, c_k will see L from the cell c_i if and only if the surface between c_i and c_j is larger than 5% of the total surface of c_j and lasts for longer than 25% of c_j cell cycle. To simulate long range actions, we then considered that the same concentration of ligand or sequestering inhibitor was present in all cell-cell interfaces with first and second order neighbors (See Fig. 5).

We optimized the parameters for this instance of our model on the embryo ASTEC-Pm8. The model performed significantly more poorly than its non diffusive counterpart. 17 over 49 cells were wrongly identified as induced (the number for the model with an action range limited to the first degree neighbor is 12), 1 known induction event was missed (the number for the model with an action range limited to the first degree neighbor is 0) as well as 1 differentially-induced cell (the number for the model with an action range limited to the first degree neighbor is 0, Table S13). In addition, the number of optimal sets of parameters significantly decreased (from 144 to only 4 combinations), suggesting that increasing the diffusion decreases the robustness of the inductions to variations in parameter values. This is further substantiated by Fig. S58, showing the number of different possible values for each model parameter, across all optimal parameter sets for both WT and diffusive conditions. All parameters except the threshold τ have only one possible optimal value in diffusive conditions. On the other hand, all parameters but ρ have, in WT conditions, more than one possible value in optimal sets. Even more importantly, the parameters directly linked to the genetic control of the inductions (the ligand concentration and affinity for its receptor, $k_f L_t$, the receptor density, $\alpha_R k_e$ and induction time) are the ones found to be the most robust in WT conditions.

9.10.2 Comparison of model results obtained with membrane-based segmentation and Voronoï tessellation

To investigate whether the precise cell geometry extracted from membrane-based segmentation was crucial for the model performance, we tested the effect of replacing it with a Voronoï tessellation of the embryo ASTEC-Pm8, built from the segmented cell barycenters. Since a Voronoï tessellation is not

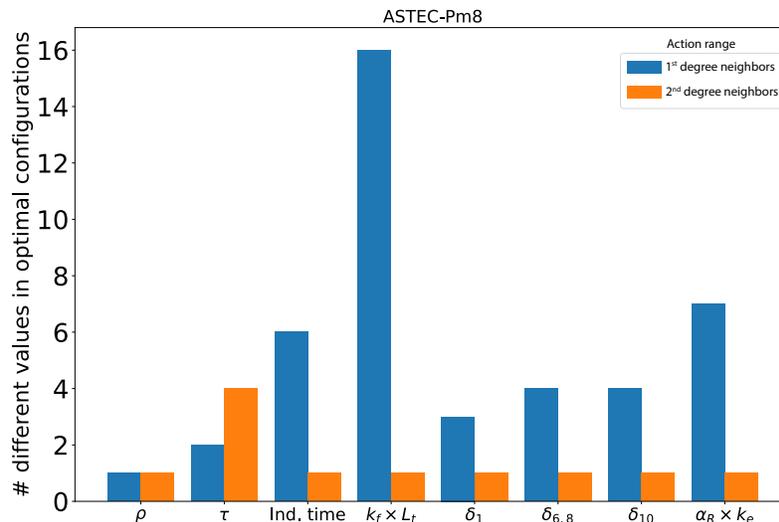


Figure S58: Number of optimal parameter values for the model run with first or second degree neighbors action range for secreted signalling proteins. The bars indicate the number of optimal values for each of the model parameters found after model training across all parameter sets without (blue bars) and with (orange bars) lateral ligand and inhibitor diffusion.

bounded in space, it is necessary to specify the boundary between the embryo and the background. In order to do so, we used the background provided by the membrane segmentation.

The model was then run on this tessellation using the optimal parameters shown in Table S12 for ASTEC-Pm8. The model only recovered 12 of 14 ground-truth differential induction events (the number for the model with the output of the ASTEC segmentation is 14) and increased the number of over-predicted differential inductions to 15 out of 49 not differentially-induced cells (the number for the model with the output of the ASTEC segmentation is 12) (Table S13).

9.10.3 Randomization of ligand gene expression pattern

To study the general sensitivity of the model to changes in expression patterns, we randomized the expression patterns and trained the model on these variant expression profiles on ASTEC-Pm1. The gene expression pattern randomization was achieved by assigning a new cell pair (at the same stage) to every cell pair/ligand expression couple. The number of expressing cells for each ligand was therefore kept constant. 200 such randomized expression patterns were generated. The model was then trained as before for each randomized gene expression pattern. On average, the model recovered 3.2 of 14 ground-truth differential induction events (14 for the model without randomization) and over predicted differential inductions for 28.9 over 49 not differentially-induced cells (12 for the model without randomization) (Table S13).

9.10.4 Constant number of receptors at cell interfaces

The model uses the geometrical information to compute the number of total receptors at the interface between cells. Its output indicates that this total number of accessible receptors provides sufficient quantitative information to explain known inductions. To assess the sensitivity of the model to the precise

geometry of cell interactions, we simulated a situation where the number of receptors at all cell-cell interfaces is identical, which amounts to only taking the topology of the embryo’s cell-cell interactions into consideration. To do so, we changed the computation of \bar{E}^* . We removed the dependency of \bar{E}^* to the cell-cell contact area in Eq. (9.22). Therefore Eq. (9.23):

$$A = \sum_{j \in N_i(L)} A_{ij}$$

becomes:

$$A = \sum_{j \in N_i(L)} \tilde{A} \tag{9.46}$$

Where \tilde{A} is the median of the areas of contact in the embryo ($\tilde{A} = 2192$ voxels, $= 197 \mu\text{m}^2$ for our embryo between time points 1 and 131). After training with these parameters, the model only predicted 4 of the 14 ground-truth induction events, and overpredicted differential induction in 7 of 49 not differentially-induced cells (Table S13).

9.10.5 Ephrin knock-out

Experimental Ephrin receptor inhibition in ascidian embryos leads to the hyperactivation of the ERK pathway (96) causing the loss of differential cell inductions of the daughters of mother cells A6.2 and A6.4 (81), and A6.3 (98). Unlike these inductions, differential mesenchyme induction in *Halocynthia* (B6.2 and B6.4 mother cells) results from the polarization of the mother cell by FGF, without antagonism from the ectoderm (32).

Inhibition of the Ephrin pathway was achieved by setting the expression of Ephrins to 0. The model was then rerun with the WT set of parameters. Under these conditions, the model only retrieved 5 of 14 differential inductions of the ground truth for cells/ligands couples. The 9 lost differential inductions all involve the ERK pathway and include the A6.2, A6.3 and A6.4 precursors whose differential inductions are also lost in response to experimental inhibition of Ephrin. Differential induction of the mesenchyme precursor B6.4 was the only preserved ERK pathway differential induction, while B6.2 induction was lost, suggesting a possible difference in the specification mechanisms of B6.4 between stolidobranchs (*Halocynthia*) and phlebobranchs (*Ciona*, *Phallusia*). As expected from the repressive action of Ephrin on the ERK pathway, differential inductions were lost as a result of the simultaneous induction of both sister cells by FGF (Fig. S59, Table S13).

10 Biological validation of the model output using live single cells measurements of ERK activity at the 64-cell stage (Supplement to Figure panels 5B)

To qualitatively compare the amount of intra-cellular effector E^* computed by our model with single-cell live measurements of FGF/Ras/ERK signaling activities, we used the ERK-Kinase Translocation Reporter, a synthetic real-time biosensor, which translates ERK activity into a nucleo-cytoplasmic shuttling event (46). We made the assumption that the amount of Ras-GTP at the plasma membrane, computed by our model, is monotonously related to ERK activity, measured by the ERK-KTR sensor.

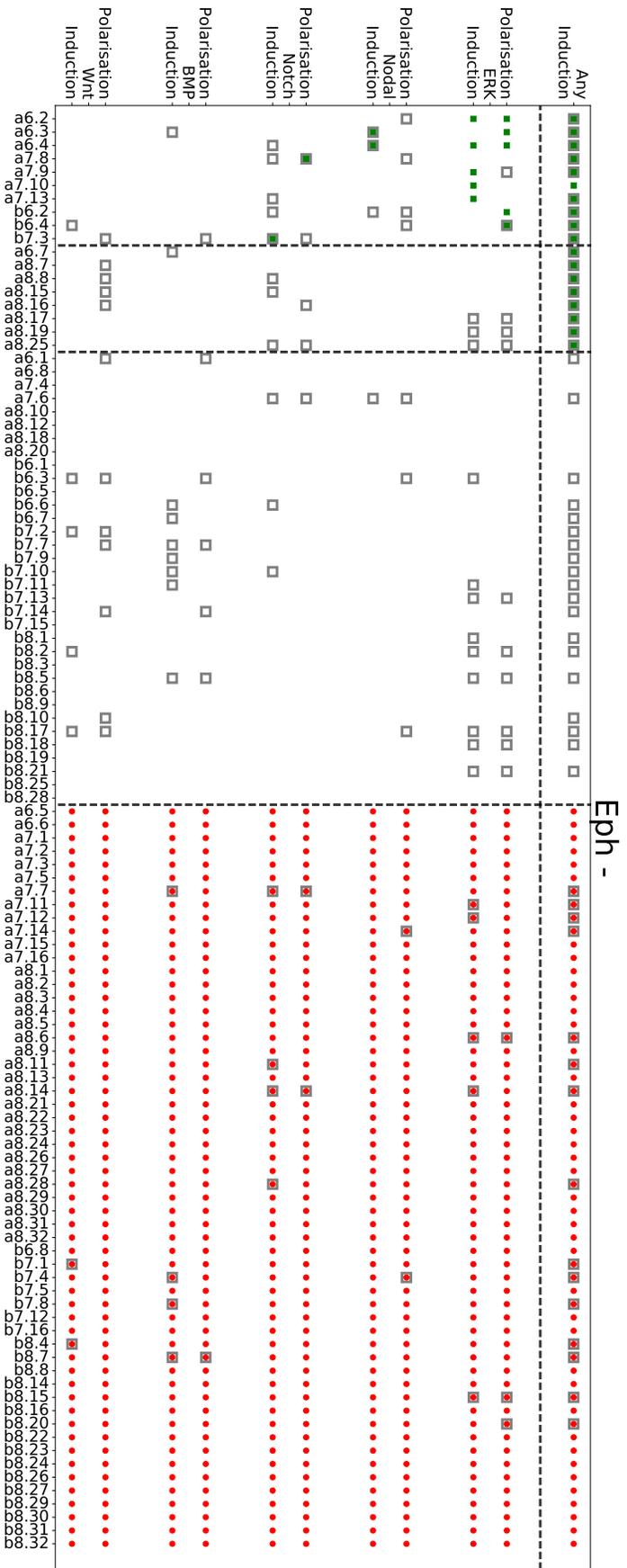


Figure S59: Results of the model run when Ephrin receptor signaling is computationally inhibited.

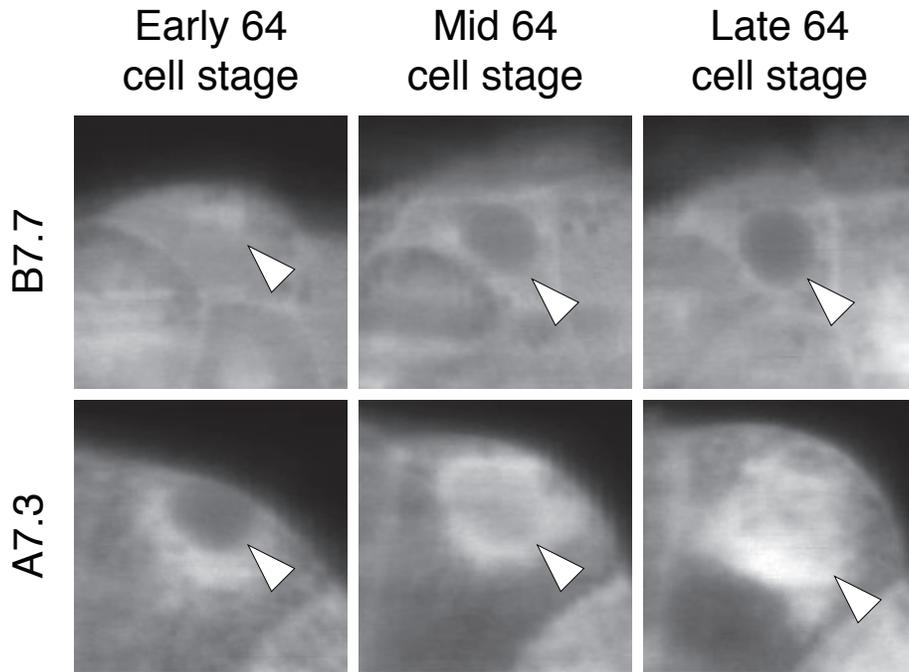


Figure S60: Examples of dynamic ERK-KTR nucleo-cytoplasmic shuttling in two example cells The figure shows the dynamic ERK-KTR nucleo-cytoplasmic shuttling in two example cells at the 64-cell stage in ASTEC-Pm13. The ERK-KTR sensor is progressively exported out of the nucleus of the B7.7 mesenchyme precursor as the cell cycle progresses, indicative of increased ERK signalling. The A7.3 notochord precursor shows the opposite behavior, consistent with the notion that the notochord fate results from the polarization of the A6.3 mother cell (80).

10.1 KTR signal analysis

10.1.1 Processing of two-channel PH-tdTomato/ERK-KTR-Clover image datasets

At each time point, four two-channels (PH-tdTomato and ERK-KTR-Clover) acquisitions are acquired. Fusion is first performed on PH-tdTomato channel (see section 2.2): resulting computed transformations as well as image weights are used to fuse the ERK-KTR-Clover channel. This results in two-channel fusion images.

To extract surface of contact areas, the fused PH-tdTomato channel was segmented using ASTEC and manually named according to the Conklin nomenclature. The fused ERK-KTR channel was then used to infer ERK activity levels in each named cell.

10.1.2 Quantification strategy for the ERK-KTR signal

As the ERK-KTR reporter is exported out of the nucleus upon phosphorylation of the reporter by active ERK kinase, we used Fiji/ImageJ (99) to measure the mean ERK-KTR fluorescence intensity in a manually-placed disk enclosed in the equatorial plane of each nucleus (we verified that the precise position and size of the disk did not significantly affect the measurements). For each embryo, we measured the ERK-KTR in all nuclei at 5 time points, spaced by 4-6 minutes. These measures were made blindly without knowledge of the cell identities. The cell naming from the segmentation was then associated with the ERK-KTR values thus obtained, revealing highly similar values in matching bilaterally

symmetric cells.

To compare the values obtained in different embryos, which may not have received the exact same amount of synthetic ERK-KTR-Clover mRNA, we normalized each nuclear intensity by the mean ERK-KTR fluorescence intensity within the whole embryo at each developmental time analyzed. These mean intensities were automatically computed as the average of the intensity in the ERK-KTR channel located within the boundary of the embryo, given by the cell segmentation of the membrane channel. Attempts to normalize the signal by the total cytoplasmic intensity of each cell gave results inconsistent with known inductions, suggesting that only part of the highly heterogeneous cytoplasmic ERK-KTR can shuttle to and from the nucleus.

10.1.3 Time registration of ERK-KTR signals in different embryos

The precise time points at which the ERK-KTR measurements were performed may differ between embryos. To compare the dynamics and the magnitude of normalised ERK-KTR signals in different embryos, a non-linear time registration of ERK-KTR signal was performed. Having measured the nuclear KTR signal for all cells in each embryo at five distinct time points, the problem consisted in identifying the temporal relation between measurements in different embryos. To this end, we employed a multi-step iterative process. At first, assuming all N_e embryos were measured at exactly the same five equally spaced timepoints 1, 2, 3, 4, 5, we calculated the average KTR signal per cell c per timepoint t as

$$\eta^{(0)}(c, t) = \frac{1}{N_e} \sum_{i=1}^{N_e} \nu_i^{(0)}(c, t), \quad (10.1)$$

where $\nu_i^{(0)}(c, t)$ is the KTR value measured for cell c at time t in embryo i . Based on this, we defined the distance in KTR signal between embryo i and the average as

$$\mathcal{D}_i^{(0)} = \frac{1}{N_c} \sum_{c=1}^{N_c} \left(\frac{1}{N_t} \sum_{t=1}^{N_t} (\nu_i^{(0)}(c, t) - \eta^{(0)}(c, t))^2 \right) = \left\langle \left\langle (\nu_i^{(0)}(c, t) - \eta^{(0)}(c, t))^2 \right\rangle_t \right\rangle_c, \quad (10.2)$$

where N_t is the number of measured timepoints per cell and N_c the number of cells in the embryo, $\langle \cdot \rangle_t$ stands for the time average for each cell and $\langle \cdot \rangle_c$ stands for the average over all cells.

The five equally spaced timepoints $t_n = n, n = 1, \dots, 5$ will serve as a reference time template for inter-embryonic comparisons of KTR signal.

Lifting now the assumption that all embryos were measured at the exact same timepoints, the set of KTR signal measurements on embryo i becomes a function of five ordered parameters $\tau_{i,1} < \tau_{i,2} < \tau_{i,3} < \tau_{i,4} < \tau_{i,5}$, corresponding at the actual time instants at which embryo i has been measured. By assuming a quasi-linear time evolution of the KTR signal between any two consecutive time instants $\tau_{i,n}, \tau_{i,n+1}$, the estimated KTR values $\nu_i^{\text{est}}(c, t_n | \tau_{i,1}, \dots, \tau_{i,5})$ per embryo and per cell at the five reference timepoints $t_n = n, n = 1, \dots, 5$ is obtained by linear interpolation between the two time instants $\tau_{i,n}, \tau_{i,n+1}$ closest to t_n , and is thus a function of the five time parameters $\tau_n, n = 1, \dots, 5$.

This also means that both the average KTR signal, Eq. (10.1), and the KTR distance between each embryo and the average, Eq. (10.2), become functions of all τ s:

$$\eta(c, t | \{\tau\}) = \frac{1}{N_e} \sum_{i=1}^{N_e} \nu_i^{\text{est}}(c, t | \tau_{i,1}, \dots, \tau_{i,5}), \quad (10.3)$$

$$\mathcal{D}_i(\{\tau\}) = \frac{1}{N_c} \sum_{c=1}^{N_c} \left(\frac{1}{N_t} \sum_{t=1}^{N_t} (\nu_i^{\text{est}}(c, t | \tau_{i,1}, \dots, \tau_{i,5}) - \eta(c, t | \{\tau\}))^2 \right), \quad (10.4)$$

where $\{\tau\}$ is a short-hand notation for $\{\tau_{1,1}, \dots, \tau_{1,5}, \tau_{2,1}, \dots, \tau_{2,5}, \dots, \tau_{N_e,5}\}$.

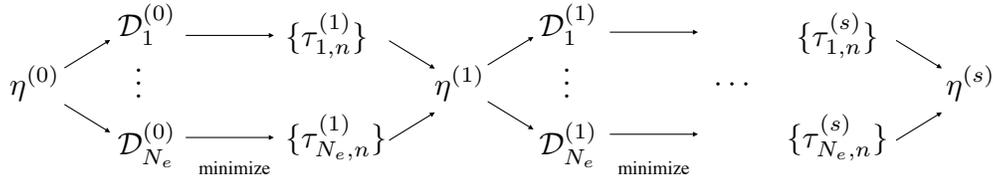


Figure S61: Schematic flow of the iterative process of time registration of KTR signal.

In order to estimate the τ parameters for each embryo, starting from the zero-th order values $\mathcal{D}_i^{(0)}$, we performed a non-linear numerical optimisation by iteratively minimizing the function $\mathcal{D}_i(\tau_1, \dots, \tau_5)$ in Eq. (10.4) to find the approximation of the τ s at each step s . These approximations, one per embryo, are then used to calculate the next estimation of the average $\eta^{(s)}$ at step s from Eq. (10.3). This procedure was performed until $\max |\eta^{(s)} - \eta^{(s-1)}| \leq 10^{-2}$, with the max performed over all cells and timepoints. This procedure is schematically represented in Fig. S61.

After convergence we obtained a reliable estimation of both the average KTR signal $\eta_{\text{conv}}(c, t_n)$ and the signal for each embryo $v_i^{\text{conv}}(c, t_n)$, $t_n = n, n = 1, \dots, 5$ per cell and per reference timepoint. This process brings all embryos onto the same time scaffold and allows for meaningful comparison of their KTR signal at corresponding timepoints.

Note that, mutant or experimentally manipulated embryos can be included in the pool of N_e embryos to which the analysis is applied, provided the perturbation does not affect the majority of cells: since Eq. (10.4) gives the average agreement per embryo, if only a small fraction of cells is impacted by the perturbation, the procedure can still be applied with relatively high confidence, as localised differences will be averaged out of the distance function. Interestingly, this procedure can efficiently single out the cells affected by the perturbation as those farthest away from the average after time registration.

10.1.4 The subcellular localization of ERK-KTR correlates with ERK activity in ascidian embryos

To confirm that the ERK-KTR-mClover sensor initially developed for mammalian cells (46) was also a reporter of ERK activity in ascidian embryos, and that our measurement strategy was suitable, we first compared the pattern of ERK-KTR nuclear accumulation with known ERK activity in ascidians. We focused our analysis at the 64-cell stage and used our census of known FGF-induced cells, which includes the progeny of 5 vegetal mother cells giving rise to differentially-induced daughters (A6.2, A6.3, A6.4, B6.2, B6.4) (Table S11) and that of the A6.1 endoderm precursor (98).

Analysis of the temporally-registered patterns of inferred ERK activity revealed a high level of reproducibility between embryos in all cells, and a clean separation of the cells known to be induced from uninduced cells (Fig. S62). Consistent with previous work in *Halocynthia roretzi* (100), the B7.1 and B7.2 posterior head endoderm progenitors strongly activate the ERK pathway (Fig. S62).

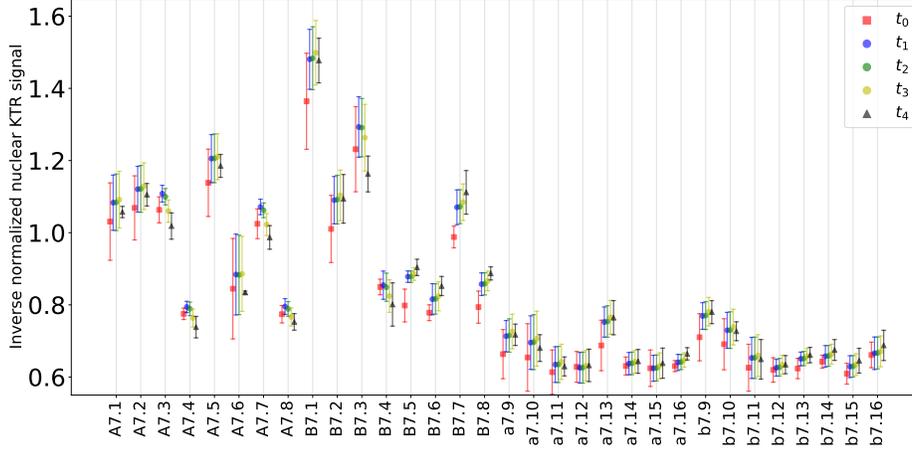


Figure S62: Average time-registered inverse normalized ERK-KTR value for wild-type embryos. The figure shows, for each cell at the 64-cell stage, the average inverse normalized value of nuclear KTR, obtained after time registration of 5 wild-type embryos. Each cell at the 64-cell stage was measured at five distinct time points (shown in different colors after time registration). Each point corresponds to the average value after time registration, with its standard deviation indicated by vertical bars. Time registration was performed according to the scheme of Fig. S61.

Consistent with the proposal that ERK activation in early ascidian embryos mostly results from the activation of the Ras/Raf/MEK cascade, pharmacological inhibition of MEK activity from the early 16-cell stage strongly reduced the activation of ERK activation (Fig. S63).

10.2 Comparison of ERK-KTR signal to model predictions

To compare the predictions of the differential induction model with the experimentally-inferred activation of ERK, we ran the differential induction model on these embryos. We first determined a *best-consensus* model parametrization. As detailed in Table S12, the optimal values for the model parameters obtained after extensive parameter-space exploration for three wild-type embryos have a narrow yet non-zero variance. In addition, more than one optimal parameter configuration was found after each parameter exploration. To determine an average parametrization of the model we looked, among all optimal configurations found for the three wild-type embryos of Table S12, for the most similar ones. To this end, each parameter configuration p for embryo $i = \text{ASTE}C - Pm1, \text{ASTE}C - Pm5, \text{ASTE}C - Pm8$ was translated into a n -dimensional vector V_{p^i} after normalising the values of each parameter in the range $[0, 1]$. We then selected, among all possible optimal configurations for the three embryos, the three configurations $q^1 = q^{\text{ASTE}C - Pm1}, q^5 = q^{\text{ASTE}C - Pm5}, q^8 = q^{\text{ASTE}C - Pm8}$ that minimize $d(V_{q^1}, V_{q^5}) + d(V_{q^1}, V_{q^8}) + d(V_{q^5}, V_{q^8})$, $d(a, b)$ being the euclidean distance between two vectors a and b . We finally defined the *best-consensus* model parameterization by taking, for each parameter, the average of its values in q^1, q^5 and q^8 . This produced the set of values given in Table S12, that were used as parameter values when running our model on embryos for which KTR signal was measured.

We then performed a correlation analysis between the registered ERK-KTR values and the model predictions of the state of activation of RasGTP, assumed to be a monotonic function of ERK_{Model}^* . Since both ERK_{Model}^* and ν_i^{CONV} are information given for each cell, we quantified how much the two

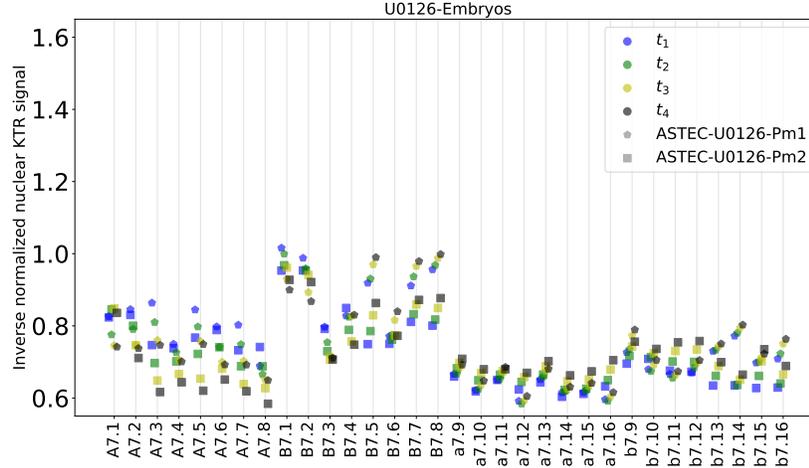


Figure S63: Time-registered inverse normalized ERK-KTR value for two embryos treated with the U0126 MEK inhibitor from the early 16-cell stage. The figure shows, for each cell at the 64-cell stage of two U0126-MEK-inhibitor-treated embryos, the time-registered inverse normalized ERK-KTR nuclear value. In both embryos and within the 64-cell stage and for each cell, the nuclear KTR was measured at 4 distinct time points (shown in different colors after time registration). Time registration was performed between the two U0126-MEK-inhibitor-treated datasets according to the scheme of Fig. S61. Note the high level of reproducibility of the measurements between embryos.

sets of data are monotonic functions of one another by computing the Spearman correlation between $ERK_{Model}^*(c)$ and $\nu_i^{conv}(c)$ as functions of embryonic cells.

11 Analysis of the effect of changes in the cell-cell contact areas (Supplementary information to Figure 5B-D)

11.1 Production, naming and analysis of half embryos

To analyze the effects of changes in the surface of contact to FGF- or Ephrin-expressing cells, we halved embryos micro-injected with mRNA for PH-tdTomato and ERK-KTR-Clover at the 2-cell stage with a thin glass filament. Embryos were imaged in the MuViSPIM for both channels (488nm and 561nm) from the 16-cell stage onwards at 18°C for 151 consecutive time points, separated by 2 minutes intervals. The membrane and KTR channels were fused, segmented, manually curated as before. Due to the perturbation, naming was more difficult than in whole embryos, and proceeded iteratively.

Overall, we considered that the halving procedure does not globally alter fate specification processes or cell identities and chose the naming compatible with the smallest number of respecified cells. Animal and vegetal hemispheres were first identified thanks to the earlier cleavage of vegetal cells at the 24- and 44-cell stages. The antero-posterior axis was then identified by looking for anterior cells in the animal and vegetal hemisphere seeding a lineage consistent with a6.5 (anterior neural plate) and with A6.2 and A6.4 (notochord and tail nerve cord identity). The identity of these cells was further confirmed by the analysis of the KTR signal and by the analysis of their relative positions (a6.5 is expected to contact A6.2 and A6.4). These cells were used as anchors to identify the halving plane of the embryo and name

the other cells following the Conklin rule: two cells sharing a mother have consecutive indices, the lower index corresponding to the cell closest to the contact point between the vegetal-most cells and the halving plane.

Having identified each cell, we measured the nuclear KTR signal (see above) at the 32- and 64-cell stages in each cell at approximately 50% of its cell cycle (Fig. 5B, C, S65 and S64). We also ran the model with the parameters of Table reftab:mod-param-ktr and compared the results to ERK activation values inferred from KTR nuclear signal measurements (Fig. 5B). Changes in fate specification processes were inferred from the comparative analysis of cell lineage trees within each analysed embryo or half embryo.

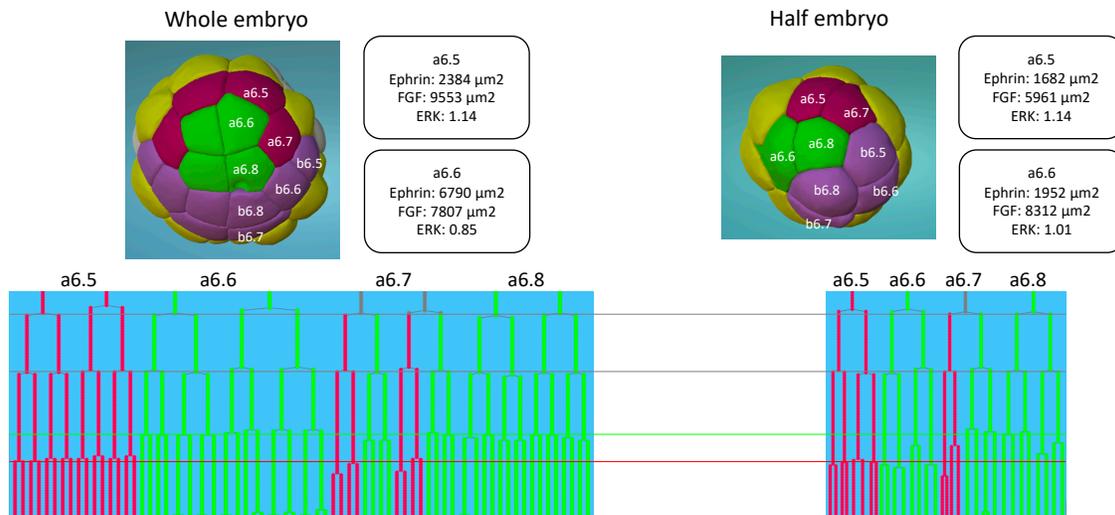


Figure S64: Effect of halving embryos on the neighborhood, ERK signalling intensity and cell lineage of the a6.5 and a6.6 blastomeres. View of *ASTECPm13* and *ASTEChalfPm1* at the 32-cell stage. Green cells are future head epidermis, red future neural cells, mauve b-line ectoderm and yellow FGF-expressing vegetal cells. Bottom panels: a-line cell lineage trees originating from the indicated cells. Grey lineages correspond to cells not yet fate restricted, green branches correspond in Whole embryos to future head epidermis, red to future neural cells. The horizontal lines indicate the timing of division of neural (red line), epidermal (green line) and undifferentiated (grey line). Note the characteristic delay in the 3rd cleavage of the neural precursors compared to epidermal precursors, and the difference in the cell lineage tree seeded by a6.6 in the whole embryo (epidermal) and in the half embryo (neural).

11.2 Response of the model to *in silico* changes in areas of cell-cell contacts

To test the robustness of inductions to changes in surfaces of contact or ligand expression levels, we looked for differential inductions/polarizations that were operating close to the induction thresholds imposed by surfaces of contact.

To identify cells operating close to surface of contact induction/polarization thresholds, we ran the model with altered surfaces of contact, keeping the set of wild type parameters, including wild-type ϵ_{min} and ϵ_{max} , the induction thresholds. We modified each surface of contact, for every cells, by $\pm x\%$ of their original size ($x \in [5, 10, \dots, 50]$). We then looked for cells which changed their induction and/or polarization status for at least one pathway.

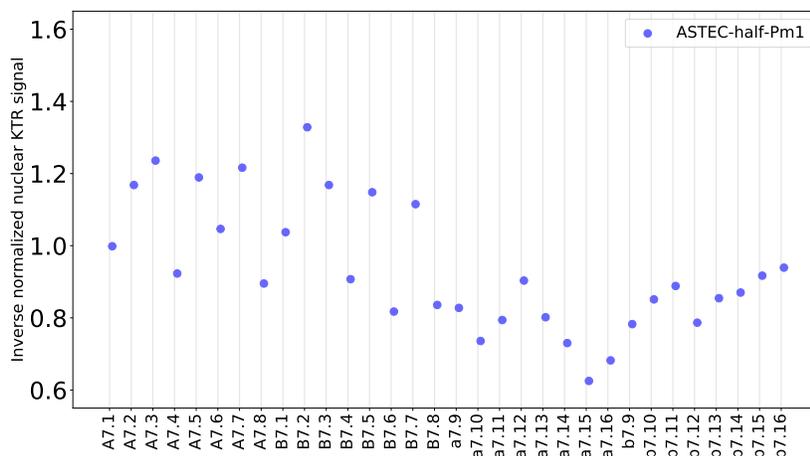


Figure S65: ERK activation in the half embryo ASTEC-half-Pm1. The figure shows the inverse normalized ERK-KTR nuclear value at the mid-64-cell stage for each cell of the ASTEC-half-Pm1 half embryo. Note that ERK signaling in A7.6 is higher than in the A7.1 and B7.1 endoderm precursors.

Because the ERK pathway works in a competitive manner (opposite effect of FGF and Ephrin on Ras activation), altering all surfaces of contact by the same percentage results in altering in similar proportions both the exposition to FGF and Ephrin. A cell having its surfaces of contact changed according to that scheme will therefore only see little changes for the ERK pathway, the increase in FGF (resp. decrease) being counterbalanced by the similar increase of Ephrin (resp. decrease). Therefore, we only changed by $\pm x\%$ the surface of contact of each cell to FGF (resp. Ephrin) without altering the surface of contact of the cells to Ephrin (resp. FGF). Then, we looked for cells that had their ERK induction and/or polarization status changed from a change of surface of contact to FGF (resp. Ephrin).

We identified cells which were close to induction/polarization thresholds as cells which geometry were constrained. The geometry of a cell was considered constrained if the cell had its induction/polarization status changed by a $\pm 30\%$ change in its surfaces of contact with a cell interface with free ligand. Fig. S66 shows the cells constrained for the 112 cell-stage, Fig. 6A shows the cells constrained for the 64 cell-stage.

11.3 Comparison of the effect of changing areas of contact vs ligand expression levels

The same overall strategy as above was used to identify cells operating close to ligand expression level (L_T) thresholds. The model was run with altered $k_f.L_T$, keeping the set of optimal wild type parameters, including wild-type ϵ_{min} and ϵ_{max} . We changed $k_f.L_T$ by $\pm x\%$ ($x \in [5, 10, \dots, 50]$). As before, we looked for cells that had their induction and/or polarization status changed. We labeled a cell as sensitive to a $x\%$ $k_f.L_t$ change if the cell had its induction and/or polarization status changed for at least one pathway in one of the two altered $x\%$ conditions (+ or - $x\%$ of $k_f.L_T$ change).

Fig. S67 shows the % of cells changing their induction/polarization status as a consequence of variations in areas of contacts or $k_f.L_T$. It illustrates the duality of action of ligand concentration and areas of contact between responsive and ligand-emitting cells.

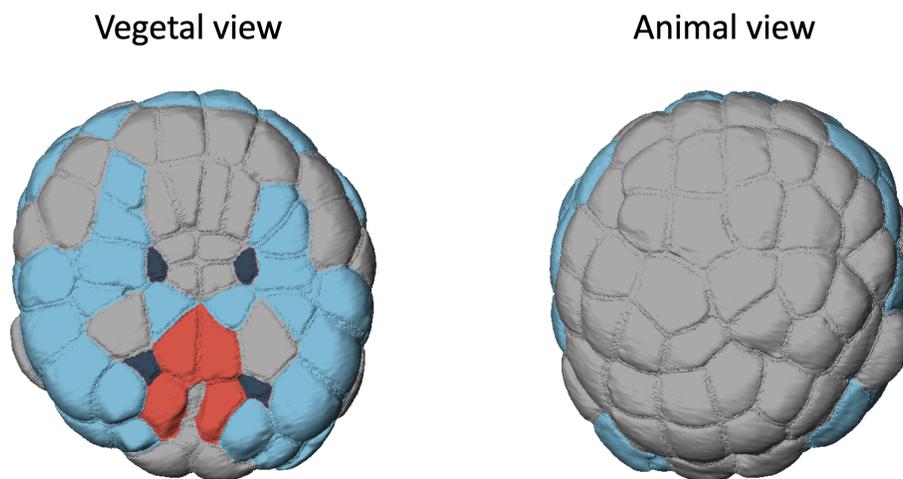


Figure S66: Sensitivity of predicted induced states to changes in the surface of contact to signalling cells. Projection onto a 112-cell stage embryo of cells, whose predicted induced state changed when exposed to a $\pm 30\%$ in silico change to their contact surface to cells expressing FGF (dark blue), ligands or inhibitors of other pathways (light blue), or both (red). Note that the induced state of most animal cells is insensitive to such surface changes. See also Fig. 6A for corresponding data data at the 64-cell stage.

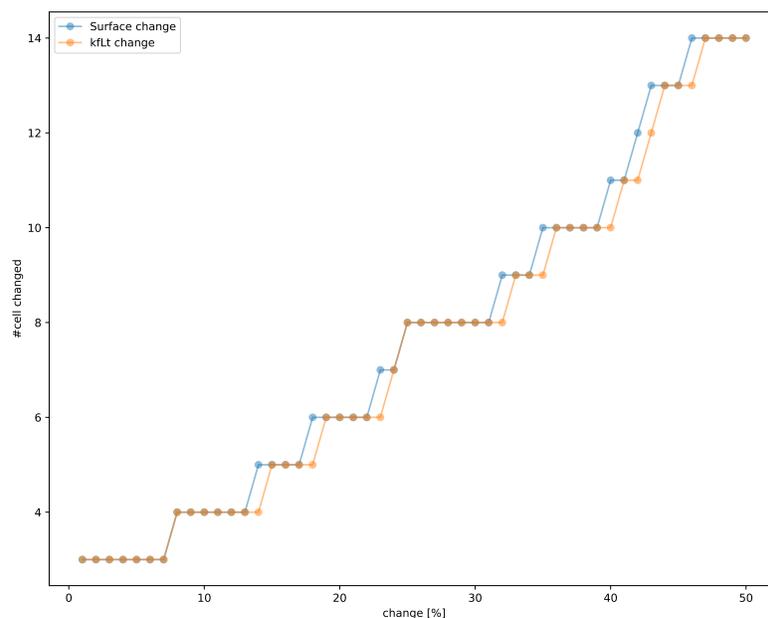


Figure S67: Comparison of the effect of altering the geometry of cell interfaces or ligand concentration on the ERK pathway induction/polarization status of cells.

12 Author contributions.

The project was initiated by Patrick Lemaire, Lars Hufnagel and Christophe Godin. The article was mostly written by Patrick Lemaire, Christophe Godin and Léo Guignard, with contributions from all authors.

Main individual contributions: Project

- Conceptualization and initiation: CG, LH, PL
- Supervision: CG, PL, GMa
- General coordination: PL

MuViSPIM imaging:

- Conceptualization: UMF, LH, PL
- Methodology (prototype): UMF
- Methodology (scaling up/Production scale): JL, UMF, (PL)
- Investigation: UMF, JL, KB

Dedicated image processing components (Image Fusion, membrane enhancement, watershed, ...) :

- Methodology/Software: GMa, GMi, (LG)

ASTEC segmentation and tracking pipeline

- Conceptualization: CG, LG, (PL)
- Methodology/Software (prototype): LG
- Methodology/Software (scaling up/Production scale): GMa, GMi, (EF, JL)
- Investigation (manual expertise/validation): UMF, (LG, JL)

Segmented series post-processing (Curation, Cell naming)

- Conceptualization: JL, BL, PL, (LG)
- Methodology/Software: BL, EF, (JL, LG)
- Data Curation: JL, (BL, LG, UMF)

Geometric data analysis

- Conceptualization: CG, LG, BL, PL, GMa
- Variability analysis: BL, LG
- Fate analysis: LG, CG, BL

Mathematical modeling of inductions:

- Conceptualization: CG, LG, PL
- Methodology/Software (prototype): LG
- Methodology/Software (scaling up): BL, LG
- Computational validation (diffusion, voronoi tessellation, ...): EF, LG, BL (PL, CG)

Experimental validation of induction model:

- Conceptualization: PL, KB
- Methodology: KB, JL
- Analysis: JL, BL, LG, (KB)

Data sharing:

- Conceptualization: GMa, PL
- Methodology: GMa, EF, (GMi)
- Data preparation: JL, GMa

13 Tables

Table S1.

***Phallusia mammillata* dataset.** Embryos ASTEC-Pm1 to -10 are wild-type embryos used to generate a geometric atlas of *Phallusia* development. Embryos ASTEC-Pm11 to -Pm15 are wild-type embryos used to estimate ERK activity during the 64-cell stage. ASTEC-half-Pm1 is a half embryo obtained after bisection along its plane of bilateral symmetry of a 2-cell embryo co-expressing PH-tdTomato and ERK-KTR-mClover. ASTEC-U0126-Pm1 and -Pm2 are embryos in which the MEK1/2 kinase was inhibited from the 16-cell stage using U0126 pharmacological inhibitor. The "Label" column indicates the fluorescent proteins expressed by each embryo. # = number. All cell and cell division numbers are extracted after segmentation and tracking, see Tables S2, S3. Naming convention for the staging: i: initial, e: early, m: mid, l: late, G: gastrula, N: neurula, TB: tailbud. (*): equivalent stage in a WT embryo. "#Cell generations" indicates the depth of the longest cell lineage in each dataset.

Embryo ID	Label	#Time points	#Cells (begin)	#Cells (end)	Dev. Stage (begin)	Dev. Stage (end)	#Cell divisions	#Cell generations
ASTEC-Pm1	PH-GFP	192	64	761	St.8 64-cell	St.17 iTB	697	7
ASTEC-Pm2	PH-GFP	95	64	574	St.8 64-cell	St.15 mN	512	6
ASTEC-Pm3	PH-citrine	90	46	325	St.7 44-cell	St.13 IG	302	5
ASTEC-Pm4	PH-citrine	90	35	272	St.6 32-cell	St.12 mG	237	5
ASTEC-Pm5	PH-citrine	90	56	331	St.7 44-cell	St.13 IG	279	5
ASTEC-Pm6	PH-citrine	90	42	311	St.7 44-cell	St.13 IG	276	5
ASTEC-Pm7	PH-citrine	80	43	220	St.7 44-cell	St.12 mG	204	5
ASTEC-Pm8	PH-citrine	88	64	397	St.8 64-cell	St.14 eN	350	6
ASTEC-Pm9	PH-citrine	63	72	300	St.8 64-cell	St.12 mG	244	5
ASTEC-Pm10	PH-tdtomato	163	47	771	St.7 44-cell	St.17 iTB	742	7
ASTEC-Pm11	PH-tdtomato KTR-clover	41	42	142	St.7 44-cell	St.11 iG	100	4
ASTEC-Pm12	PH-tdtomato KTR-clover	23	60	110	St.8 64-cell	St.11 iG	50	3
ASTEC-Pm13	PH-tdtomato KTR-clover	90	25	228	St.6 32-cell	St.12 mG	203	5
ASTEC-Pm14	PH-tdtomato KTR-clover	28	32	76	St.6 32-cell	St.9 76-cell	44	3
ASTEC-Pm15	PH-tdtomato KTR-clover	90	16	185	St.5 16-cell	St.11 eG	169	5
ASTEC-half-Pm1	PH-tdtomato KTR-clover	151	16	320	St.6 32-cell	St.16 IN(*)	314	7
ASTEC-U0126-Pm1	PH-tdtomato KTR-clover	115	16	364	St.5 16-cell	St.13 IG(*)	356	7
ASTEC-U0126-Pm2	PH-tdtomato KTR-clover	21	42	80	St.7 44-cell	St.9 76-cell(*)	38	3

Table S2.

Automatic post-corrections by the ASTEC pipeline This Table recapitulates the over-segmentation corrections made by the automatic post-correction algorithm at the end of the ASTEC pipeline. We distinguish here a cell, which exists over several time points, from a cell snapshot, which represents the cell (or a part of the cell in case of over-segmentation) at one imaged time point of its life. "#cell snapshots before post-corr." gives the number of cell snapshots across the whole movie before the post-segmentation step. "#corrected over-seg" refers to the number of over-segmentations that were corrected by fusion of all snapshots corresponding to the same cell at a given time point. Once an over-segmentation occurs, it is usually carried over several time-points, generating an artefactual lineage branch. "#corrected over-seg events" is the number of such branches that have been removed from the lineage tree during the post-correction step. "#corrected over-seg cells" is the number of cells that were corrected for over-segmentation at least once during their cell cycle (a single cell could have been over-segmented at several, non consecutive, time points, this number is hence smaller than "#corrected over-seg events"). "#cell divisions after post-corr." is the number of cell divisions detected after the post-correction step.

Embryo	#cell snapshots before post-corr.	#corrected over-seg	#corrected over-seg events	#corrected over-seg cells	#cell divisions after post-corr.
ASTEC-Pm1	74723	7720	1295	576	697
ASTEC-Pm2	23528	976	79	63	512
ASTEC-Pm3	14023	388	36	33	302
ASTEC-Pm4	12318	627	51	42	237
ASTEC-Pm5	14343	523	69	53	279
ASTEC-Pm6	12395	172	23	21	276
ASTEC-Pm7	12725	1771	132	82	204
ASTEC-Pm8	17743	406	60	56	350
ASTEC-Pm9	10775	371	46	38	244
ASTEC-Pm10	60821	903	113	91	742

Table S3.

Manual corrections and assessment of final segmentation and tracking. Following post-correction, a manual correction/curation step was implemented to correct residual over-segmentations and to detect errors left uncorrected by both automated and manual corrections (under-segmentation, missed cells, lost cell tracking). "#cell snapshots" is the number of snapshots after post-correction. "#cells" is the number of cells after manual correction. "#cell divisions" is the number of cell divisions after manual curation. "#corrected over-seg events" is the number of branches that have been manually removed from the lineage tree as a result of the fusion of the snapshots corresponding to the same cell at all over-segmentation time points (E_s errors). "#missed cell divisions" is the number of cell divisions that were never resolved (i.e. the pipeline never detected the cell division and the two sisters remained fused as a single cell). "#missed cells" is the number of cells that were never detected (typically, when a mother cell divides, the division is detected at the manual curation step but one of the two sisters is not detected, leaving a hole in the segmented embryo). "#interrupt. lineage track" is the number of E_{mi} errors: cells initially detected and which disappear prior to any cell division and before the end of the movie. *: the higher number of missed cells in ASTEC-Pm2 is due to a partial cropping of some neural plate cells during imaging.

Embryo	#cell snapshots	#cells	#cell divisions	#corrected over-seg. events	#missed cell divisions	#missed cells	#interrupt. lineage track
ASTEC-Pm1	67003	1395	697	1	0	4	1
ASTEC-Pm2	22552	1066	501	5	0	11*	2
ASTEC-Pm3	13635	604	279	21	1	1	0
ASTEC-Pm4	11691	507	237	0	1	0	0
ASTEC-Pm5	13820	608	276	7	2	0	1
ASTEC-Pm6	12223	580	269	17	1	0	0
ASTEC-Pm7	10954	391	174	30	2	0	0
ASTEC-Pm8	17337	728	332	29	0	1	1
ASTEC-Pm9	10404	530	229	29	2	2	2
ASTEC-Pm10	59918	1519	736	19	5	4	15
Total	239537	7928	3730	158	14	21	22

Table S4.

Ground truth for known asymmetric cell divisions. Cells between the 32-cell stage and the last division of an 8th generation cell. In blue the cells that were considered for this study (according to the rules defined in 7.2)

Mother cell $X_{n.p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Status
A6.2	1st Lineage, Notochord	Posterior Ventral Neural Plate	Asym
A6.3	Head Endoderm	Trunk Lateral Cell	Asym
A6.4	1st Lineage, Notochord	Undetermined	Asym
a6.7	Anterior Dorsal Neural Plate	Head Epidermis	Asym
B6.1	Head Endoderm	Undetermined	Asym
B6.2	Undetermined	1st Lineage, Tail Muscle	Asym
B6.3	Trunk Ventral Cell	Germ Line	Asym
B6.4	Mesenchyme	1st Lineage, Tail Muscle	Asym
b6.5	Undetermined	Undetermined	Asym
A7.6	Trunk Lateral Cell	Trunk Lateral Cell	Asym
A7.8	Posterior Lateral Neural Plate	Undetermined	Asym
a7.9	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a7.10	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a7.13	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym
B7.3	Mesenchyme	2nd Lineage, Notochord	Asym
b7.9	Undetermined	Tail Epidermis	Asym
b7.10	Posterior Dorsal Neural Plate	Tail Epidermis	Asym
A8.7	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym
A8.8	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym
A8.15	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym
A8.16	2nd Lineage, Tail Muscle	Posterior Lateral Neural Plate	Asym
a8.17	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a8.18	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a8.19	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a8.20	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym
a8.25	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym
a8.26	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym
B8.9	Tail Muscle	Trunk Ventral Cell	Asym
B8.10	Tail Muscle	Trunk Ventral Cell	Asym
A9.14	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym
A9.16	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym
A9.29	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym

Table S5.

Ground truth for known symmetric cell divisions II (no fate restriction events). Cells between the 32-cell stage and the last division of an 8th generation cell. In blue the cells that were considered for this study (according to the rules defined in 7.2)

Mother cell $X_{n.p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Status
A6.1	Head Endoderm	Head Endoderm	Sym
a6.6	Head Epidermis	Head Epidermis	Sym
a6.8	Head Epidermis	Head Epidermis	Sym
b6.8	Tail Epidermis	Undetermined	Sym
A7.3	1st Lineage, Notochord	1st Lineage, Notochord	Sym
A7.4	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Sym
A7.7	1st Lineage Notochord	1st Lineage, Notochord	Sym
a7.11	Head Epidermis	Head Epidermis	Sym
a7.12	Head Epidermis	Head Epidermis	Sym
a7.14	Head Epidermis	Head Epidermis	Sym
a7.15	Head Epidermis	Head Epidermis	Sym
a7.16	Head Epidermis	Head Epidermis	Sym
B7.4	1st Lineage, Tail Muscle	1st Lineage, Tail Muscle	Sym
B7.5	Trunk Ventral Cell	Trunk Ventral Cell	Sym
B7.8	1st Lineage Tail Muscle	1st Lineage Tail Muscle	Sym
b7.15	Tail Epidermis	Tail Epidermis	Sym
b7.16	Undetermined	Undetermined	Sym
A8.5	1st Lineage, Notochord	1st Lineage, Notochord	Sym
A8.6	1st Lineage, Notochord	1st Lineage, Notochord	Sym
A8.13	1st Lineage, Notochord	1st Lineage, Notochord	Sym
A8.14	1st Lineage, Notochord	1st Lineage, Notochord	Sym
a8.21	Head Epidermis	Head Epidermis	Sym
a8.22	Head Epidermis	Head Epidermis	Sym
a8.23	Head Epidermis	Head Epidermis	Sym
a8.24	Head Epidermis	Head Epidermis	Sym
a8.27	Head Epidermis	Head Epidermis	Sym
a8.28	Head Epidermis	Head Epidermis	Sym
a8.29	Head Epidermis	Head Epidermis	Sym
a8.30	Head Epidermis	Head Epidermis	Sym
a8.31	Head Epidermis	Head Epidermis	Sym
a8.32	Head Epidermis	Head Epidermis	Sym
B8.6	2nd Lineage, Notochord	2nd Lineage, Notochord	Sym
B8.7	1st Lineage, Tail Muscle	1st Lineage, Tail Muscle	Sym
B8.8	1st Lineage, Tail Muscle	1st Lineage, Tail Muscle	Sym
B8.16	1st Lineage, Tail Muscle	1st Lineage, Tail Muscle	Sym
b8.26	Tail Epidermis	Tail Epidermis	Sym
b8.29	Tail Epidermis	Tail Epidermis	Sym
b8.30	Tail Epidermis	Tail Epidermis	Sym
A9.26	1st Lineage, Notochord	1st Lineage, Notochord	Sym
a9.41-48	Head Epidermis	Head Epidermis	Sym
a9.53-62,a9.64	Head Epidermis	Head Epidermis	Sym
a9.63	Head Epidermis	Head Epidermis	Sym

Table S6.

Asymmetric division found by the classifier for known and candidate asymmetric cell divisions (fate restriction events). This Table lists known (Source ANISEED, (61)) and candidate asymmetric cell divisions from the 32-cell to the late neurula stage. In addition to cell identities and fates, it lists for each mother the average and standard deviation of the three features described previously. The number of embryos analyzed is defined by the depth of cell divisions undergone by the cell considered as explained in 8.1.1.

Mother cell $X_{n,p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Status (n of embryos)	V_i	$Lted_i$	Lt_i
A6.2	1st Lineage Notochord	Posterior Ventral Neural Plate	known (9)	1.13/0.06	0.15/0.05	1.10/0.06
A6.3	Head Endoderm	Trunk Lateral Cell	known (10)	1.07/0.06	0.19/0.04	1.80/0.58
A6.4	1st Lineage Notochord	Undetermined	known (9)	1.21/0.07	0.10/0.05	1.18/0.15
a6.7	Anterior Dorsal Neural Plate	Head Epidermis	known (10)	1.23/0.10	0.11/0.06	1.02/0.02
B6.2	Undetermined	1st Lineage Tail Muscle	known (9)	1.17/0.08	0.10/0.04	1.03/0.04
B6.3	Trunk Ventral Cell	Germ Line	known (3)	6.70/1.62	0.40/0.00	2.79/0.58
B6.4	Mesenchyme	1st Lineage Tail Muscle	known (5)	1.66/0.09	0.15/0.04	2.55/0.68
A7.6	Trunk Lateral Cell	Trunk Lateral Cell	known (5)	1.29/0.19	0.15/0.08	1.23/0.14
A7.8	Posterior Lateral Neural Plate	Undetermined	known (7)	1.08/0.07	0.13/0.08	1.02/0.02
a7.9	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	known (6)	1.28/0.11	0.06/0.05	1.06/0.03
a7.10	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	known (6)	1.37/0.12	0.08/0.04	1.04/0.02
a7.13	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	known (4)	1.51/0.21	0.11/0.01	1.14/0.03
B7.3	Mesenchyme	2nd Lineage Notochord	known (8)	3.13/1.12	0.15/0.08	1.32/0.17
b7.9	Undetermined	Tail Epidermis	known (6)	1.36/0.23	0.13/0.08	1.29/0.13
b7.10	Posterior Dorsal Neural Plate	Tail Epidermis	known (7)	1.68/0.21	0.16/0.06	1.30/0.04
A8.7	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	known (2)	2.38/0.37	0.17/0.02	1.27/0.18
A8.8	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	known (4)	2.88/0.35	0.10/0.03	1.22/0.05
A8.15	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	known (3)	1.37/0.12	0.08/0.05	1.14/0.05
a8.18	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	known (3)	1.77/0.18	0.14/0.06	1.13/0.03
a8.19	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	known (2)	1.31/0.13	0.08/0.03	1.15/0.06
a8.20	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	known (3)	1.59/0.17	0.10/0.08	1.11/0.06
a8.25	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	known (2)	3.24/0.88	0.21/0.01	1.68/0.08
A9.29	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	known (1)	1.17/0.00	0.13/0.00	1.09/0.00
A7.1	Head Endoderm	Head Endoderm	candidate (1)	1.00/0.00	0.06/0.00	1.00/0.00
A7.5	Head Endoderm	Head Endoderm	candidate (2)	1.03/0.00	0.10/0.03	1.10/0.06
B7.7	Mesenchyme	Mesenchyme	candidate (1)	11.75/0.00	0.38/0.00	1.02/0.00
b7.13	Tail Epidermis	Tail Epidermis	candidate (9)	1.56/0.22	0.09/0.05	1.03/0.02
A8.12	Trunk Lateral Cell	Trunk Lateral Cell	candidate (2)	1.08/0.08	0.07/0.01	1.06/0.00
b8.18	Tail Epidermis	Tail Epidermis	candidate (3)	1.40/0.10	0.10/0.08	1.09/0.06
b8.21	Tail Epidermis	Tail Epidermis	candidate (3)	1.55/0.09	0.15/0.11	2.01/0.78
b8.25	Tail Epidermis	Tail Epidermis	candidate (2)	1.45/0.15	0.18/0.04	1.10/0.03
b8.28	Tail Epidermis	Tail Epidermis	candidate (3)	1.62/0.49	0.18/0.06	1.17/0.10
a9.36	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	candidate (2)	1.07/0.01	0.09/0.04	1.03/0.00
B9.9	Mesenchyme	Mesenchyme	candidate (1)	1.08/0.00	0.14/0.00	1.40/0.00
A10.57	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	candidate (1)	1.00/0.00	0.18/0.00	1.03/0.00

Table S7.

Volume ratio between sister cells. This Table lists known asymmetric and symmetric cell divisions (Source ANISEED, (61)) from the 32-cell to the late neurula stage that are available within our dataset (see 7.2 for a description of how the cells were chosen). It lists cell identities and fates and the average and standard deviation of the volume ratio described previously.

Mother cell $X_{n,p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Division Symmetry (n of embryos)	V_i	Above threshold
A6.2	1st Lineage Notochord	Posterior Ventral Neural Plate	Asym (9)	1.13/0.06	False
A6.3	Head Endoderm	Trunk Lateral Cell	Asym (10)	1.07/0.06	False
A6.4	1st Lineage Notochord	Undetermined	Asym (9)	1.21/0.07	True
a6.7	Anterior Dorsal Neural Plate	Head Epidermis	Asym (10)	1.23/0.10	True
B6.1	Head Endoderm	Undetermined	Asym (7)	1.10/0.07	False
B6.2	Undetermined	1st Lineage Tail Muscle	Asym (9)	1.17/0.08	True
B6.3	Trunk Ventral Cell	Germ Line	Asym (3)	6.70/1.62	True
B6.4	Mesenchyme	1st Lineage Tail Muscle	Asym (5)	1.66/0.09	True
b6.5	Undetermined	Undetermined	Asym (9)	1.07/0.04	False
A7.6	Trunk Lateral Cell	Trunk Lateral Cell	Asym (5)	1.29/0.19	True
A7.8	Posterior Lateral Neural Plate	Undetermined	Asym (7)	1.08/0.07	False
a7.9	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (6)	1.28/0.11	True
a7.10	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (6)	1.37/0.12	True
a7.13	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (4)	1.51/0.21	True
B7.3	Mesenchyme	2nd Lineage Notochord	Asym (8)	3.13/1.12	True
b7.9	Undetermined	Tail Epidermis	Asym (6)	1.36/0.23	True
b7.10	Posterior Dorsal Neural Plate	Tail Epidermis	Asym (7)	1.68/0.21	True
A8.7	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (2)	2.38/0.37	True
A8.8	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (4)	2.88/0.35	True
A8.15	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym (3)	1.37/0.12	True
a8.17	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (2)	1.30/0.16	True
a8.18	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (3)	1.77/0.18	True
a8.19	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (2)	1.31/0.13	True
a8.20	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (3)	1.59/0.17	True
a8.25	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (2)	3.24/0.88	True
a8.26	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (2)	1.01/0.01	False
A9.14	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (2)	1.02/0.01	False
A9.29	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym (1)	1.17/0.00	True
A6.1	Head Endoderm	Head Endoderm	Sym (8)	1.07/0.05	False
a6.6	Head Epidermis	Head Epidermis	Sym (10)	1.05/0.05	False
a6.8	Head Epidermis	Head Epidermis	Sym (10)	1.13/0.09	False
b6.8	Tail Epidermis	Undetermined	Sym (10)	1.09/0.07	False
A7.3	1st Lineage Notochord	1st Lineage Notochord	Sym (4)	1.05/0.03	False
A7.4	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Sym (9)	1.03/0.03	False
A7.7	1st Lineage Notochord	1st Lineage Notochord	Sym (3)	1.11/0.03	False
a7.11	Head Epidermis	Head Epidermis	Sym (9)	1.03/0.03	False
a7.12	Head Epidermis	Head Epidermis	Sym (9)	1.05/0.04	False
a7.14	Head Epidermis	Head Epidermis	Sym (8)	1.07/0.04	False
a7.15	Head Epidermis	Head Epidermis	Sym (9)	1.05/0.06	False
a7.16	Head Epidermis	Head Epidermis	Sym (8)	1.08/0.04	False
B7.4	1st Lineage Tail Muscle	1st Lineage Tail Muscle	Sym (4)	1.02/0.01	False
b7.15	Tail Epidermis	Tail Epidermis	Sym (9)	1.05/0.05	False
b7.16	Undetermined	Undetermined	Sym (9)	1.08/0.08	False
a8.21	Head Epidermis	Head Epidermis	Sym (3)	1.15/0.10	False
a8.22	Head Epidermis	Head Epidermis	Sym (3)	1.14/0.11	False
a8.23	Head Epidermis	Head Epidermis	Sym (3)	1.06/0.04	False
a8.24	Head Epidermis	Head Epidermis	Sym (3)	1.06/0.01	False
a8.27	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.00	False
a8.28	Head Epidermis	Head Epidermis	Sym (3)	1.02/0.01	False
a8.29	Head Epidermis	Head Epidermis	Sym (3)	1.03/0.02	False
a8.30	Head Epidermis	Head Epidermis	Sym (3)	1.09/0.02	False
a8.31	Head Epidermis	Head Epidermis	Sym (3)	1.09/0.07	False
a8.32	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.01	False
b8.26	Tail Epidermis	Tail Epidermis	Sym (3)	1.05/0.05	False
b8.29	Tail Epidermis	Tail Epidermis	Sym (3)	1.03/0.02	False
b8.30	Tail Epidermis	Tail Epidermis	Sym (2)	1.06/0.01	False
A9.26	1st Lineage Notochord	1st Lineage Notochord	Sym (1)	1.10/0.00	False
a9.63	Head Epidermis	Head Epidermis	Sym (1)	1.07/0.00	False

Table S8.

Volume ratio between sister cells II. Same Table as Table S7 for sister cells with no known information about the symmetry of their division.

Mother cell $X_{n.p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Division Symmetry (n of embryos)	V_i	Above threshold
a6.5	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Unknown (10)	1.02/0.02	False
b6.6	Tail Epidermis	Tail Epidermis	Unknown (10)	1.22/0.16	True
b6.7	Tail Epidermis	Tail Epidermis	Unknown (10)	1.18/0.10	True
A7.1	Head Endoderm	Head Endoderm	Unknown (1)	1.00/0.00	False
A7.2	Head Endoderm	Head Endoderm	Unknown (3)	1.02/0.01	False
A7.5	Head Endoderm	Head Endoderm	Unknown (2)	1.03/0.00	False
B7.1	Head Endoderm	Head Endoderm	Unknown (2)	1.19/0.04	True
B7.2	Head Endoderm	1st Endodermal Lineage	Unknown (2)	1.16/0.08	False
B7.7	Mesenchyme	Mesenchyme	Unknown (1)	11.75/0.00	True
b7.11	Tail Epidermis	Tail Epidermis	Unknown (9)	1.42/0.25	True
b7.12	Tail Epidermis	Tail Epidermis	Unknown (9)	1.13/0.08	False
b7.13	Tail Epidermis	Tail Epidermis	Unknown (9)	1.56/0.22	True
b7.14	Tail Epidermis	Tail Epidermis	Unknown (10)	1.36/0.13	True
A8.3	Head Endoderm	Head Endoderm	Unknown (1)	1.25/0.00	True
A8.11	Trunk Lateral Cell	Trunk Lateral Cell	Unknown (1)	1.03/0.00	False
A8.12	Trunk Lateral Cell	Trunk Lateral Cell	Unknown (2)	1.08/0.08	False
B8.1	Head Endoderm	Head Endoderm	Unknown (1)	1.07/0.00	False
B8.4	1st Endodermal Lineage	1st Endodermal Lineage	Unknown (1)	1.09/0.00	False
B8.5	Mesenchyme	Mesenchyme	Unknown (2)	1.10/0.07	False
b8.18	Tail Epidermis	Tail Epidermis	Unknown (3)	1.40/0.10	True
b8.19	Posterior Dorsal Neural Plate	Posterior Dorsal Neural Plate	Unknown (1)	1.10/0.00	False
b8.20	Tail Epidermis	Tail Epidermis	Unknown (3)	1.10/0.02	False
b8.21	Tail Epidermis	Tail Epidermis	Unknown (3)	1.55/0.09	True
b8.22	Tail Epidermis	Tail Epidermis	Unknown (3)	1.05/0.04	False
b8.23	Tail Epidermis	Tail Epidermis	Unknown (3)	1.02/0.02	False
b8.24	Tail Epidermis	Tail Epidermis	Unknown (3)	1.07/0.04	False
b8.25	Tail Epidermis	Tail Epidermis	Unknown (2)	1.45/0.15	True
b8.27	Tail Epidermis	Tail Epidermis	Unknown (3)	1.01/0.01	False
b8.28	Tail Epidermis	Tail Epidermis	Unknown (3)	1.62/0.49	True
b8.31	Tail Epidermis	Tail Epidermis	Unknown (3)	1.07/0.02	False
b8.32	Tail Epidermis	Tail Epidermis	Unknown (3)	1.04/0.02	False
a9.36	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Unknown (2)	1.07/0.01	False
a9.51	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Unknown (1)	1.11/0.00	False
B9.9	Mesenchyme	Mesenchyme	Unknown (1)	1.08/0.00	False
B9.10	Mesenchyme	Mesenchyme	Unknown (1)	1.02/0.00	False
b9.35	Tail Epidermis	Tail Epidermis	Unknown (1)	1.03/0.00	False
b9.42	Tail Epidermis	Tail Epidermis	Unknown (1)	1.07/0.00	False
b9.52	Tail Epidermis	Tail Epidermis	Unknown (1)	1.06/0.00	False
b9.55	Tail Epidermis	Tail Epidermis	Unknown (1)	1.15/0.00	False
A10.57	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Unknown (1)	1.00/0.00	False

Table S9.

Cell cycle length ratio between sister cells. This Table lists known asymmetric and symmetric cell divisions (Source ANISEED, (61)) from the 32-cell to the late neurula stage that are available within our dataset (see 7.2 for a description of how the cells were chosen). It lists cell identities and fates and the average and standard deviation of the volume ratio described previously.

Mother cell $X_{n,p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Division Symmetry (n of embryos)	L_i	Above threshold
A6.2	1st Lineage Notochord	Posterior Ventral Neural Plate	Asym (9)	1.10/0.06	False
A6.3	Head Endoderm	Trunk Lateral Cell	Asym (10)	1.80/0.58	True
A6.4	1st Lineage Notochord	Undetermined	Asym (9)	1.18/0.15	True
a6.7	Anterior Dorsal Neural Plate	Head Epidermis	Asym (10)	1.02/0.02	False
B6.1	Head Endoderm	Undetermined	Asym (7)	1.07/0.07	False
B6.2	Undetermined	1st Lineage Tail Muscle	Asym (9)	1.03/0.04	False
B6.3	Trunk Ventral Cell	Germ Line	Asym (3)	2.79/0.58	True
B6.4	Mesenchyme	1st Lineage Tail Muscle	Asym (5)	2.55/0.68	True
b6.5	Undetermined	Undetermined	Asym (9)	1.03/0.05	False
A7.6	Trunk Lateral Cell	Trunk Lateral Cell	Asym (5)	1.23/0.14	True
A7.8	Posterior Lateral Neural Plate	Undetermined	Asym (7)	1.02/0.02	False
a7.9	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (6)	1.06/0.03	False
a7.10	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (6)	1.04/0.02	False
a7.13	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (4)	1.14/0.03	False
B7.3	Mesenchyme	2nd Lineage Notochord	Asym (8)	1.32/0.17	True
b7.9	Undetermined	Tail Epidermis	Asym (6)	1.29/0.13	True
b7.10	Posterior Dorsal Neural Plate	Tail Epidermis	Asym (7)	1.30/0.04	True
A8.7	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (2)	1.27/0.18	True
A8.8	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (4)	1.22/0.05	True
A8.15	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym (3)	1.14/0.05	False
a8.17	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (2)	1.22/0.08	True
a8.18	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (3)	1.13/0.03	False
a8.19	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (2)	1.15/0.06	False
a8.20	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	Asym (3)	1.11/0.06	False
a8.25	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (2)	1.68/0.08	True
a8.26	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	Asym (2)	1.03/0.00	False
A9.14	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Asym (2)	1.02/0.00	False
A9.29	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	Asym (1)	1.09/0.00	False
A6.1	Head Endoderm	Head Endoderm	Sym (8)	1.05/0.04	False
a6.6	Head Epidermis	Head Epidermis	Sym (10)	1.05/0.09	False
a6.8	Head Epidermis	Head Epidermis	Sym (10)	1.00/0.00	False
b6.8	Tail Epidermis	Undetermined	Sym (10)	1.02/0.04	False
A7.3	1st Lineage Notochord	1st Lineage Notochord	Sym (4)	1.03/0.02	False
A7.4	Posterior Ventral Neural Plate	Posterior Ventral Neural Plate	Sym (9)	1.02/0.02	False
A7.7	1st Lineage Notochord	1st Lineage Notochord	Sym (3)	1.05/0.02	False
a7.11	Head Epidermis	Head Epidermis	Sym (9)	1.02/0.04	False
a7.12	Head Epidermis	Head Epidermis	Sym (9)	1.01/0.02	False
a7.14	Head Epidermis	Head Epidermis	Sym (8)	1.01/0.02	False
a7.15	Head Epidermis	Head Epidermis	Sym (9)	1.00/0.00	False
a7.16	Head Epidermis	Head Epidermis	Sym (8)	1.00/0.00	False
B7.4	1st Lineage Tail Muscle	1st Lineage Tail Muscle	Sym (4)	1.01/0.02	False
b7.15	Tail Epidermis	Tail Epidermis	Sym (9)	1.01/0.02	False
b7.16	Undetermined	Undetermined	Sym (9)	1.02/0.02	False
a8.21	Head Epidermis	Head Epidermis	Sym (3)	1.02/0.01	False
a8.22	Head Epidermis	Head Epidermis	Sym (3)	1.03/0.00	False
a8.23	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.02	False
a8.24	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.01	False
a8.27	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.02	False
a8.28	Head Epidermis	Head Epidermis	Sym (3)	1.00/0.00	False
a8.29	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.01	False
a8.30	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.01	False
a8.31	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.02	False
a8.32	Head Epidermis	Head Epidermis	Sym (3)	1.01/0.01	False
b8.26	Tail Epidermis	Tail Epidermis	Sym (3)	1.03/0.02	False
b8.29	Tail Epidermis	Tail Epidermis	Sym (3)	1.01/0.01	False
b8.30	Tail Epidermis	Tail Epidermis	Sym (2)	1.00/0.00	False
A9.26	1st Lineage Notochord	1st Lineage Notochord	Sym (1)	1.00/0.00	False
a9.63	Head Epidermis	Head Epidermis	Sym (1)	1.00/0.00	False

Table S10.

Cell cycle length ratio between sister cells II. Same Table as Table S9 for sister cells with no known information about the symmetry of their division.

Mother cell $X_{n.p}$	Fate daughter $X(n+1).(2p-1)$	Fate daughter $X(n+1).2p$	Division Symmetry (n of embryos)	V_i	Above threshold
a6.5	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	grey (10)	1.01/0.02	False
b6.6	Tail Epidermis	Tail Epidermis	grey (10)	1.03/0.05	False
b6.7	Tail Epidermis	Tail Epidermis	grey (10)	1.02/0.02	False
A7.1	Head Endoderm	Head Endoderm	grey (1)	1.00/0.00	False
A7.2	Head Endoderm	Head Endoderm	grey (3)	1.16/0.16	False
A7.5	Head Endoderm	Head Endoderm	grey (2)	1.10/0.06	False
B7.1	Head Endoderm	Head Endoderm	grey (2)	1.07/0.02	False
B7.2	Head Endoderm	1st Endodermal Lineage	grey (2)	1.02/0.01	False
B7.7	Mesenchyme	Mesenchyme	grey (1)	1.02/0.00	False
b7.11	Tail Epidermis	Tail Epidermis	grey (9)	1.00/0.00	False
b7.12	Tail Epidermis	Tail Epidermis	grey (9)	1.01/0.02	False
b7.13	Tail Epidermis	Tail Epidermis	grey (9)	1.03/0.02	False
b7.14	Tail Epidermis	Tail Epidermis	grey (10)	1.04/0.04	False
A8.3	Head Endoderm	Head Endoderm	grey (1)	1.02/0.00	False
A8.11	Trunk Lateral Cell	Trunk Lateral Cell	grey (1)	1.02/0.00	False
A8.12	Trunk Lateral Cell	Trunk Lateral Cell	grey (2)	1.06/0.00	False
B8.1	Head Endoderm	Head Endoderm	grey (1)	1.07/0.00	False
B8.4	1st Endodermal Lineage	1st Endodermal Lineage	grey (1)	1.02/0.00	False
B8.5	Mesenchyme	Mesenchyme	grey (2)	1.02/0.02	False
b8.18	Tail Epidermis	Tail Epidermis	grey (3)	1.09/0.06	False
b8.19	Posterior Dorsal Neural Plate	Posterior Dorsal Neural Plate	grey (1)	1.06/0.00	False
b8.20	Tail Epidermis	Tail Epidermis	grey (3)	1.02/0.03	False
b8.21	Tail Epidermis	Tail Epidermis	grey (3)	2.01/0.78	True
b8.22	Tail Epidermis	Tail Epidermis	grey (3)	1.01/0.02	False
b8.23	Tail Epidermis	Tail Epidermis	grey (3)	1.02/0.02	False
b8.24	Tail Epidermis	Tail Epidermis	grey (3)	1.01/0.02	False
b8.25	Tail Epidermis	Tail Epidermis	grey (2)	1.10/0.03	False
b8.27	Tail Epidermis	Tail Epidermis	grey (3)	1.02/0.02	False
b8.28	Tail Epidermis	Tail Epidermis	grey (3)	1.17/0.10	False
b8.31	Tail Epidermis	Tail Epidermis	grey (3)	1.00/0.00	False
b8.32	Tail Epidermis	Tail Epidermis	grey (3)	1.01/0.02	False
a9.36	Anterior Ventral Neural Plate	Anterior Ventral Neural Plate	grey (2)	1.03/0.00	False
a9.51	Anterior Dorsal Neural Plate	Anterior Dorsal Neural Plate	grey (1)	1.05/0.00	False
B9.9	Mesenchyme	Mesenchyme	grey (1)	1.40/0.00	True
B9.10	Mesenchyme	Mesenchyme	grey (1)	1.02/0.00	False
b9.35	Tail Epidermis	Tail Epidermis	grey (1)	1.09/0.00	False
b9.42	Tail Epidermis	Tail Epidermis	grey (1)	1.00/0.00	False
b9.52	Tail Epidermis	Tail Epidermis	grey (1)	1.03/0.00	False
b9.55	Tail Epidermis	Tail Epidermis	grey (1)	1.03/0.00	False
A10.57	Posterior Lateral Neural Plate	Posterior Lateral Neural Plate	grey (1)	1.03/0.00	False

Table S11.

List of experimentally-determined differential sister inductions and mother polarizations leading to differential fates acquisition by sister cells. Top Table: Differential inductions where both inducers, modes of inductions and stage of induction are known. Bottom Table: Inductions for which the mechanism and timing of induction are unknown.

Cell pair	Mother polarization(St.)	Daughter induction (St.)	Reference
A6.2	ERK (St. 6)	ERK (St. 8)	(81); (101)
A6.3	ERK (St. 6)	Nodal (St. 8)	(98)
A6.4	ERK (St. 6)	ERK, Nodal (St. 8)	(81); (101); (102)
B6.2	ERK (St. 6)	•	(103)
B6.4	ERK (St. 6)	•	(104)
A7.8	Notch (St. 8)	•	(78)
a7.9	•	ERK (St. 10)	(105)
a7.10	•	ERK (St. 10)	(105)
a7.13	•	ERK (St. 10)	(105)
B7.3	•	Notch (St. 10)	(76)

Cell pair	Mother polarization or Daughter induction	Reference
a6.7	Bmp	(96)
A8.7	ERK	(78)
A8.8	ERK	(78)
A8.15	ERK	(78)
A8.16	ERK	(78)
a8.17	ERK	(105)
a8.19	ERK	(105)
a8.25	ERK	(105)

Table S12.
Comprehensive list of model parameters with optimal values.

Embryo name	ρ	τ	Time of induction (min)	$k_f.L_T$ (s^{-1})	δ_1	$\delta_{6,8}$	δ_{10}	$\alpha_R.k_e$ ($\mu\text{m}^{-2}s^{-1}$)
ASTEC-Pm1	1.8	44%	8	2×10^{-4}	10	4×10^{-2}	0.32	10^{-6}
ASTEC-Pm5	2.2	46%	12	6.3×10^{-6}	31.6	3×10^{-3}	3×10^{-2}	6×10^{-7}
ASTEC-Pm8	2.4	46%	12	2.2×10^{-6}	100	9×10^{-4}	3×10^{-2}	6×10^{-7}
Best Consensus	2.1	45%	12	4.4×10^{-6}	47.2	2.3×10^{-3}	3×10^{-2}	9×10^{-7}

Table S13.

Model predictions and parametrization Comparison of model predictions obtained for the three digitized embryos ASTEC-Pm1, ASTEC-Pm5 and ASTEC-Pm8 under different model assumptions and geometries. Column (1): % of known differential induction events correctly predicted, including ligand identity ($n = 14$); column (2): % of over-predicted differential inductions ($n = 49$ sister cell pairs not differentially induced); column (3) % of known differentially-induced cells for which the induction pathway may be unknown ($n = 18$). The computation of the metrics for columns (1), (2) and (3) were sometimes irrelevant, in these cases we entered “N/A” (not applicable) in the respective columns. (4) Parametrization. “Trained” indicates that the model predictions reflect the best fit that can be obtained by training the model on the specific assumption tested. ASTEC-Pm1 indicate that the model was run using an optimal parameter set for ASTEC-Pm1 WT conditions (first row). If not indicated otherwise, the secreted protein range is limited to first-order neighbors.

Embryo/Perturbation	(1) True Positives (with pathway)	(2) False Positives	(3) True Positives	(4) Parametrization
ASTEC-Pm1/Wild type	100%	22%	100%	Trained
ASTEC-Pm5/Wild type	86%	22%	100%	Trained
ASTEC-Pm8/Wild type	100%	24%	100%	Trained
ASTEC-Pm1/Gene pattern expression randomization ($N = 100$)	25%	52%	80%	Trained
ASTEC-Pm1/Constant number of receptors at each cell-cell interface	29%	13%	83%	Trained
ASTEC-Pm1/Ephrin virtual knock-out	29%	27%	94%	ASTEC-Pm1
ASTEC-Pm1/4x uniform scaling of surfaces	100%	22%	100%	Trained
ASTEC-Pm1/Local random variations in cell-cell contact areas ($\bar{n} = 20\%$, $N = 100$)	77%	22%	89%	ASTEC-Pm1
ASTEC-Pm8/Voronoi tessellation	86%	31%	83%	Trained
ASTEC-Pm8/WT Range of secreted proteins extended to 2nd-order neighbors	92%	35%	94%	Trained
ASTEC-Pm5 w/ ASTEC-Pm1 parameters	86%	33%	67%	ASTEC-Pm1
ASTEC-Pm8 w/ ASTEC-Pm1 parameters	71%	24%	78%	ASTEC-Pm1
ASTEC-Pm1/ $\pm 30\%$ change in surface to cell expressing FGF	N/A	N/A	N/A	ASTEC-Pm1
ASTEC-Pm1/ $\pm 30\%$ change in surface to cell expressing ligands or inhibitors of other pathways	N/A	N/A	N/A	ASTEC-Pm1
ASTEC-half-Pm1	N/A	N/A	N/A	Best consensus parameters (S12)

14 Access to datasets and software.

14.1 Software and tutorials

Github repositories

The software, standard parameters files and tutorials can be found in the Github repository: <https://github.com/astec-segmentation>.

- astec-2019-published is the repository dedicated to this paper with a fixed and autonomous (full version, python and C codes) ASTEC software package. This include all the codes and libraries necessary to install the software with examples as standard parameters files and a tutorial to test it. A complete documentation is also provided to guide the users.
- astec is the repository for for the ongoing development, enhancement and storage the package development process. Only the python codes are included in this repository. The standards parameter files or tutorial provided in the astec-2019-published repository may have to be adapted to more recent versions of the code.
- Standard parameters. We provide a standard example parameter file that can be initially used to setup each ASTEC step on either the test set we provide (see below) or to initiate a project on personal data.
- Tutorial and test dataset. We provide a tutorial to test each step of ASTEC, a set of 20 timepoints from ASTEC-Pm1 with a downsized resolution to $1\mu\text{m}^3$. The ASTEC pipeline on this dataset can be run a standard workstation (4 threads; < 1 mn for the fusion and segmentation of a timepoint).

A Jupyter notebook providing examples on how to use the outputs of ASTEC can be found in the Github repository: <https://github.com/leoguignard/ASTEC-examples>.

14.2 Movies

Movie S1. 3D Rendering of intensity

3D rendering of an intensity image of developing ASTEC-Pm1 after fusion. Left: vegetal view, right: lateral view. Anterior is to the top.

Movie S2. 3D Rendering of intensity for 10 embryos

3D rendering of the intensity images of the ten ASTEC embryos. Vegetal view. Anterior is to the top.

Movie S3. ASTEC segmentation, clonal view

Vegetal view of the ASTEC-Pm1 segmented embryo. Color code arbitrary. Note the shape of clones produced from individual 64-cell precursors. Anterior is to the top.

Movie S4. ASTEC segmentation, Fate map view, 10 embryos

Vegetal view of the ASTEC-Pm1 to -Pm10 segmented and fate colored *Phallusia mammillata* embryos. Anterior is to the top. Dark grey: undetermined; blue: A-line head endoderm; light green: B-line head endoderm; yellow: endodermal strand; light pink: trunk ventral cells and germ line; light blue: trunk lateral cells; white: primary notochord; orange: secondary notochord; red: primary tail muscle; dark green;

mesenchyme; light pink : germ line; pink, grey, dark red, beige: neural plate derivatives; dark violet: anterior epidermis; light green midline tail epidermis; very light blue: medio-lateral tail epidermis; dark beige: lateral tail epidermis.

14.3 Datasets

Imaging, segmentations and geometric data The imaging, segmentation or geometric informations for the following ASTEC datasets can be downloaded from Figshare (<https://figshare.com/s/765d4361d1b073beedd5>).

- Raw images: The four views for some time points of the movie Astec-Pm1 are shared to test the fusion algorithms or evaluate the quality of the acquisition obtained with the MuViSPIM light-sheet microscope. The format of the images is .hdf5 files that can be easily read with an ImageJ/FIJI plug-in.
- Fused images: complete sequence of fused images are shared for ASTEC-Pm1 to -Pm15, ASTEC-U0126-Pm1 and Pm2, ASTEC-half-Pm1.
- Segmented images: complete sequence of segmented images (voxelic representation) are provided for the entire ASTEC dataset.
- Meshed segmented images: a meshed version of the segmented images (.obj format), which can be uploaded into the interactive Morphonet web tool, is provided for the complete dataset.
- Geometric properties: Properties measured after spatial registration of the movies are shared in both .pkl and .xml format for the complete dataset.

Note : For each embryo, all files are compressed into a single archive file (tar.gz). A tool provided in the astec-2019-published Github repository can be used to convert .inr format files into the more common .mha files.

Data S1. Astec-Pm1, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-17:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8223890>

Content:

- Rawdata for 20 distinct times points are provided in the raw.tar.gz.
- The four angle fusion images obtained using the ASTEC pipeline are shared in .inr format (fuse.tar.gz archive file).
- Whole cells were segmented using the ASTEC pipeline and shared in .inr format (post.tar.gz archive file).
- Surface meshes (.obj) provided in the mesh.tar.gz archive were produced using the VTK library and MeshLab.
- Individual cell geometric properties are shared in two different formats, .xml and .pkl (properties.tar.gz archive file).

Data S2. Astec-Pm2 Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-15:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235443>

Content:

- The four angle fusion images obtained using the ASTEC pipeline are shared in .inr format (fuse.tar.gz archive file).
- Whole cells were segmented using the ASTEC pipeline and shared in .inr format (post.tar.gz archive file).
- Surface meshes (.obj) provided in the mesh.tar.gz archive were produced using the VTK library and MeshLab.
- Individual cell geometric properties are shared in two different formats, .xml and .pkl (properties.tar.gz archive file).

Data S3. Astec-Pm3, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-13:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235449>

Content: (see **Data S2**)

Data S4. Astec-Pm4, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 6-12:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235455>

Content: (see **Data S2**)

Data S5. Astec-Pm5, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-13:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235458>

Content: (see **Data S2**)

Data S6. Astec-Pm6, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-13:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235470>

Content: (see **Data S2**)

Data S7. Astec-Pm7, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-12:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235473>

Content: (see **Data S2**)

Data S8. Astec-Pm8, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-14:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235479>

Content: (see **Data S2**)

Data S9. Astec-Pm9, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-12):

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235482>

Content: (see **Data S2**)

Data S10. Astec-Pm10, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-17:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235485>

Content: (see **Data S2**)

Data S11. Astec-Pm11, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 7-11):

<https://figshare.com/s/765d4361d1b073beedd5#/articles/8235485>

Content:

- The four angle fusion images obtained using the ASTEC pipeline are shared in .inr format (fuse.tar.gz archive file).
- Whole cells were segmented using the ASTEC pipeline and shared in .inr format (post.tar.gz archive file).
- Individual cell geometric properties are shared in two different formats, .xml and .pkl (properties.tar.gz archive file).

Data S12. Astec-Pm12, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-10:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11301857>

Content: (see **Data S11**)

Data S13. Astec-Pm13, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-10:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11309315>

Content: (see **Data S11**)

Data S14. Astec-Pm14, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-10:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11309381>

Content: (see **Data S11**)

Data S15. Astec-Pm15, Wild type

Phallusia mammillata embryo, live SPIM imaging, stages 8-10:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11309381>

Content: (see **Data S11**)

Data S16. Astec-U0126-Pm1, U0126 treated

Phallusia mammillata embryo, live SPIM imaging, stages 5-13:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11307431>

Content: (see **Data S11**)

Data S17. Astec-U0126-Pm2, U0126 treated

Phallusia mammillata embryo, live SPIM imaging, stages 7-9:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11307293>

Content: (see **Data S11**)

Data S18. Astec-half-Pm1, half embryo

Cut at 2-cell stage, half *Phallusia mammillata* embryo, live SPIM imaging, stages 6-16:

<https://figshare.com/s/765d4361d1b073beedd5#/articles/11309570>

Content: (see **Data S11**)

Interactive exploration of datasets through the Morphonet morphogenetic browser. Meshed and annotated ASTEC-Pm1 to -Pm10 datasets can be interactively explored through Morphonet (<http://www.morphonet.org/>) (login: phallusia; password: ASTEC2019) upon selection of the *Phallusia mammillata* species.

References and Notes

1. A. Pierre, J. Sallé, M. Wühr, N. Minc, Generic theoretical models to predict division patterns of cleaving embryos. *Dev. Cell* **39**, 667–682 (2016). [doi:10.1016/j.devcel.2016.11.018](https://doi.org/10.1016/j.devcel.2016.11.018) [Medline](#)
2. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983). [doi:10.1016/0012-1606\(83\)90201-4](https://doi.org/10.1016/0012-1606(83)90201-4) [Medline](#)
3. P. Lemaire, Evolutionary crossroads in developmental biology: The tunicates. *Development* **138**, 2143–2152 (2011). [doi:10.1242/dev.048975](https://doi.org/10.1242/dev.048975) [Medline](#)
4. F. Delsuc, H. Brinkmann, D. Chourrout, H. Philippe, Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006). [doi:10.1038/nature04336](https://doi.org/10.1038/nature04336) [Medline](#)
5. B. Goldstein, On the evolution of early development in the Nematoda. *Philos. Trans. R. Soc. London Ser. B* **356**, 1521–1531 (2001). [doi:10.1098/rstb.2001.0977](https://doi.org/10.1098/rstb.2001.0977) [Medline](#)
6. F. Delsuc, H. Philippe, G. Tsagkogeorga, P. Simion, M.-K. Tilak, X. Turon, S. López-Legentil, J. Piette, P. Lemaire, E. J. P. Douzery, A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* **16**, 39 (2018). [doi:10.1186/s12915-018-0499-2](https://doi.org/10.1186/s12915-018-0499-2) [Medline](#)
7. A. Stolfi, E. K. Lowe, C. Racioppi, F. Ristoratore, C. T. Brown, B. J. Swalla, L. Christiaen, Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. *eLife* **3**, e03728 (2014). [doi:10.7554/eLife.03728](https://doi.org/10.7554/eLife.03728) [Medline](#)
8. H. Nishida, Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121**, 526–541 (1987). [doi:10.1016/0012-1606\(87\)90188-6](https://doi.org/10.1016/0012-1606(87)90188-6) [Medline](#)
9. D. A. Voronov, Y. V. Panchin, Cell lineage in marine nematode *Enoplus brevis*. *Development* **125**, 143–150 (1998). [Medline](#)
10. H. Nishida, Specification of embryonic axis and mosaic development in ascidians. *Dev. Dyn.* **233**, 1177–1193 (2005). [doi:10.1002/dvdy.20469](https://doi.org/10.1002/dvdy.20469) [Medline](#)
11. P. Lemaire, Unfolding a chordate developmental program, one cell at a time: Invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. *Dev. Biol.* **332**, 48–60 (2009). [doi:10.1016/j.ydbio.2009.05.540](https://doi.org/10.1016/j.ydbio.2009.05.540) [Medline](#)
12. O. Tassy, F. Daian, C. Hudson, V. Bertrand, P. Lemaire, A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. *Curr. Biol.* **16**, 345–358 (2006). [doi:10.1016/j.cub.2005.12.044](https://doi.org/10.1016/j.cub.2005.12.044) [Medline](#)
13. G. de Medeiros, N. Norlin, S. Gunther, M. Albert, L. Panavaite, U.-M. Fiuza, F. Peri, T. Hiiragi, U. Krzic, L. Hufnagel, Confocal multiview light-sheet microscopy. *Nat. Commun.* **6**, 8881 (2015). [doi:10.1038/ncomms9881](https://doi.org/10.1038/ncomms9881) [Medline](#)
14. V. Ulman, M. Maška, K. E. G. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, I. Smal, K. Rohr, J. Jaldén, H. M. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S.-Y. Cho, A. C. Dufour, J.-C. Olivo-Marin, C. C. Reyes-Aldasoro, J. A. Solis-Lemus, R. Bensch, T. Brox, J. Stegmaier, R. Mikut, S. Wolf, F. A. Hamprecht, T. Esteves, P. Quelhas, Ö. Demirel, L. Malmström, F. Jug, P. Tomancak, E. Meijering, A. Muñoz-Barrutia, M. Kozubek, C. Ortiz-de-Solorzano, An objective comparison of cell-tracking algorithms. *Nat. Methods* **14**, 1141–1152 (2017). [doi:10.1038/nmeth.4473](https://doi.org/10.1038/nmeth.4473) [Medline](#)

15. E. Faure, T. Savy, B. Rizzi, C. Melani, O. Stašová, D. Fabrèges, R. Špir, M. Hammons, R. Čunderlík, G. Recher, B. Lombardot, L. Duloquin, I. Colin, J. Kollár, S. Desnoullez, P. Affaticati, B. Maury, A. Boyreau, J.-Y. Nief, P. Calvat, P. Vernier, M. Frain, G. Lutfalla, Y. Kergosien, P. Suret, M. Remešiková, R. Doursat, A. Sarti, K. Mikula, N. Peyriéras, P. Bourguine, A workflow to process 3D+time microscopy images of developing organisms and reconstruct their cell lineage. *Nat. Commun.* **7**, 8674 (2016). [doi:10.1038/ncomms9674](https://doi.org/10.1038/ncomms9674) [Medline](#)
16. J. Stegmaier, F. Amat, W. C. Lemon, K. McDole, Y. Wan, G. Teodoro, R. Mikut, P. J. Keller, Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Dev. Cell* **36**, 225–240 (2016). [doi:10.1016/j.devcel.2015.12.028](https://doi.org/10.1016/j.devcel.2015.12.028) [Medline](#)
17. Y. Azuma, S. Onami, Biologically constrained optimization based cell membrane segmentation in *C. elegans* embryos. *BMC Bioinformatics* **18**, 307 (2017). [doi:10.1186/s12859-017-1717-6](https://doi.org/10.1186/s12859-017-1717-6) [Medline](#)
18. P. Villoutreix, J. Delile, B. Rizzi, L. Duloquin, T. Savy, P. Bourguine, R. Doursat, N. Peyriéras, An integrated modelling framework from cells to organism based on a cohort of digital embryos. *Sci. Rep.* **6**, 37438 (2016). [doi:10.1038/srep37438](https://doi.org/10.1038/srep37438) [Medline](#)
19. M. A. Luengo-Oroz, D. Pastor-Escuredo, C. Castro-Gonzalez, E. Faure, T. Savy, B. Lombardot, J. L. Rubio-Guivernau, L. Duloquin, M. J. Ledesma-Carbayo, P. Bourguine, N. Peyriéras, A. Santos, 3D+t morphological processing: Applications to embryogenesis image analysis. *IEEE Trans. Image Process.* **21**, 3518–3530 (2012). [doi:10.1109/TIP.2012.2197007](https://doi.org/10.1109/TIP.2012.2197007) [Medline](#)
20. R. Fernandez, P. Das, V. Mirabet, E. Moscardi, J. Traas, J.-L. Verdeil, G. Malandain, C. Godin, Imaging plant growth in 4D: Robust tissue reconstruction and lineaging at cell resolution. *Nat. Methods* **7**, 547–553 (2010). [doi:10.1038/nmeth.1472](https://doi.org/10.1038/nmeth.1472) [Medline](#)
21. B. Leggio, J. Laussu, A. Carlier, C. Godin, P. Lemaire, E. Faure, *MorphoNet*: An interactive online morphological browser to explore complex multi-scale data. *Nat. Commun.* **10**, 2812 (2019). [doi:10.1038/s41467-019-10668-1](https://doi.org/10.1038/s41467-019-10668-1) [Medline](#)
22. W. R. Jeffery, Programmed cell death in the ascidian embryo: Modulation by FoxA5 and Manx and roles in the evolution of larval development. *Mech. Dev.* **118**, 111–124 (2002). [doi:10.1016/S0925-4773\(02\)00236-8](https://doi.org/10.1016/S0925-4773(02)00236-8) [Medline](#)
23. X. Li, Z. Zhao, W. Xu, R. Fan, L. Xiao, X. Ma, Z. Du, Systems properties and spatiotemporal regulation of cell position variability during embryogenesis. *Cell Rep.* **26**, 313–321.e7 (2019). [doi:10.1016/j.celrep.2018.12.052](https://doi.org/10.1016/j.celrep.2018.12.052) [Medline](#)
24. E. M. Leffler, K. Bullaughey, D. R. Matute, W. K. Meyer, L. Ségurel, A. Venkat, P. Andolfatto, M. Przeworski, Revisiting an old riddle: What determines genetic diversity levels within species? *PLOS Biol.* **10**, e1001388 (2012). [doi:10.1371/journal.pbio.1001388](https://doi.org/10.1371/journal.pbio.1001388) [Medline](#)
25. X. Morin, Y. Bellaïche, Mitotic spindle orientation in asymmetric and symmetric cell divisions during animal development. *Dev. Cell* **21**, 102–119 (2011). [doi:10.1016/j.devcel.2011.06.012](https://doi.org/10.1016/j.devcel.2011.06.012) [Medline](#)
26. Y. Arata, H. Takagi, Y. Sako, H. Sawa, Power law relationship between cell cycle duration and cell volume in the early embryonic development of *Caenorhabditis elegans*. *Front. Physiol.* **5**, 529 (2015). [doi:10.3389/fphys.2014.00529](https://doi.org/10.3389/fphys.2014.00529) [Medline](#)
27. S. Strome, Determination of cleavage planes. *Cell* **72**, 3–6 (1993). [doi:10.1016/0092-8674\(93\)90041-N](https://doi.org/10.1016/0092-8674(93)90041-N) [Medline](#)

28. A. Madgwick, M. S. Magri, C. Dantec, D. Gailly, U.-M. Fiuza, L. Guignard, S. Hettinger, J. L. Gomez-Skarmeta, P. Lemaire, Evolution of embryonic *cis*-regulatory landscapes between divergent *Phallusia* and *Ciona* ascidians. *Dev. Biol.* **448**, 71–87 (2019). [doi:10.1016/j.ydbio.2019.01.003](https://doi.org/10.1016/j.ydbio.2019.01.003) [Medline](#)
29. U.-M. Fiuza, T. Negishi, A. Rouan, H. Yasuo, P. Lemaire, A Nodal/Eph signalling relay drives the transition from apical constriction to apico-basal shortening in ascidian endoderm invagination. bioRxiv 418988 [Preprint]. 1 December 2019. <https://doi.org/10.1101/418988>.
30. J. Dardaillon, D. Dauga, P. Simion, E. Faure, T. A. Onuma, M. B. DeBiasse, A. Louis, K. R. Nitta, M. Naville, L. Besnardeau, W. Reeves, K. Wang, M. Fagotto, M. Guéroult-Bellone, S. Fujiwara, R. Dumollard, M. Veeman, J.-N. Volff, H. Roest Crolius, E. Douzery, J. F. Ryan, B. Davidson, H. Nishida, C. Dantec, P. Lemaire, ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. *Nucleic Acids Res.* **48**, D668–D675 (2019). [doi:10.1093/nar/gkz955](https://doi.org/10.1093/nar/gkz955) [Medline](#)
31. G. J. Kim, A. Yamada, H. Nishida, An FGF signal from endoderm and localized factors in the posterior-vegetal egg cytoplasm pattern the mesodermal tissues in the ascidian embryo. *Development* **127**, 2853–2862 (2000). [Medline](#)
32. G. J. Kim, G. Kumano, H. Nishida, Cell fate polarization in ascidian mesenchyme/muscle precursors by directed FGF signaling and role for an additional ectodermal FGF antagonizing signal in notochord/nerve cord precursors. *Development* **134**, 1509–1518 (2007). [doi:10.1242/dev.02825](https://doi.org/10.1242/dev.02825) [Medline](#)
33. C. de la Cova, R. Townley, S. Regot, I. Greenwald, A real-time biosensor for ERK activity reveals signaling dynamics during *C. elegans* cell fate specification. *Dev. Cell* **42**, 542–553.e4 (2017). [doi:10.1016/j.devcel.2017.07.014](https://doi.org/10.1016/j.devcel.2017.07.014) [Medline](#)
34. C. Cao, L. A. Lemaire, W. Wang, P. H. Yoon, Y. A. Choi, L. R. Parsons, J. C. Matese, W. Wang, M. Levine, K. Chen, Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019). [doi:10.1038/s41586-019-1385-y](https://doi.org/10.1038/s41586-019-1385-y) [Medline](#)
35. M. Nitzan, N. Karaikos, N. Friedman, N. Rajewsky, Gene expression cartography. *Nature* **576**, 132–137 (2019). [doi:10.1038/s41586-019-1773-3](https://doi.org/10.1038/s41586-019-1773-3) [Medline](#)
36. K. Sugimura, P.-F. Lenne, F. Graner, Measuring forces and stresses in situ in living tissues. *Development* **143**, 186–196 (2016). [doi:10.1242/dev.119776](https://doi.org/10.1242/dev.119776) [Medline](#)
37. K. Sherrard, F. Robin, P. Lemaire, E. Munro, Sequential activation of apical and basolateral contractility drives ascidian endoderm invagination. *Curr. Biol.* **20**, 1499–1510 (2010). [doi:10.1016/j.cub.2010.06.075](https://doi.org/10.1016/j.cub.2010.06.075) [Medline](#)
38. A. Sagner, J. Briscoe, Morphogen interpretation: Concentration, time, competence, and signaling dynamics. *WIREs Dev. Biol.* **6**, e271 (2017). [doi:10.1002/wdev.271](https://doi.org/10.1002/wdev.271) [Medline](#)
39. O. Shaya, U. Binshtok, M. Hersch, D. Rivkin, S. Weinreb, L. Amir-Zilberstein, B. Khamaisi, O. Oppenheim, R. A. Desai, R. J. Goodyear, G. P. Richardson, C. S. Chen, D. Sprinzak, Cell-cell contact area affects notch signaling and notch-dependent patterning. *Dev. Cell* **40**, 505–511.e6 (2017). [doi:10.1016/j.devcel.2017.02.009](https://doi.org/10.1016/j.devcel.2017.02.009) [Medline](#)
40. F. Xiong, A. R. Tentner, P. Huang, A. Gelas, K. R. Mosaliganti, L. Souhait, N. Rannou, I. A. Swinburne, N. D. Obholzer, P. D. Cowgill, A. F. Schier, S. G. Megason, Specified neural progenitors sort to form sharp domains after noisy Shh signaling. *Cell* **153**, 550–561 (2013). [doi:10.1016/j.cell.2013.03.023](https://doi.org/10.1016/j.cell.2013.03.023) [Medline](#)

41. C. Sardet, A. McDougall, H. Yasuo, J. Chenevert, G. Pruliere, R. Dumollard, C. Hudson, C. Hebras, N. Le Nguyen, A. Paix, Embryological methods in ascidians: The Villefranche-sur-Mer protocols. *Methods Mol. Biol.* **770**, 365–400 (2011). [doi:10.1007/978-1-61779-210-6_14](https://doi.org/10.1007/978-1-61779-210-6_14) [Medline](#)
42. C. Hudson, S. Darras, D. Caillol, H. Yasuo, P. Lemaire, A conserved role for the MEK signalling pathway in neural tissue specification and posteriorisation in the invertebrate chordate, the ascidian *Ciona intestinalis*. *Development* **130**, 147–159 (2003). [doi:10.1242/dev.00200](https://doi.org/10.1242/dev.00200) [Medline](#)
43. G. J. Kim, H. Nishida, Role of the FGF and MEK signaling pathway in the ascidian embryo. *Dev. Growth Differ.* **43**, 521–533 (2001). [doi:10.1046/j.1440-169X.2001.00594.x](https://doi.org/10.1046/j.1440-169X.2001.00594.x) [Medline](#)
44. E. R. Liman, J. Tytgat, P. Hess, Subunit stoichiometry of a mammalian K⁺ channel determined by construction of multimeric cDNAs. *Neuron* **9**, 861–871 (1992). [doi:10.1016/0896-6273\(92\)90239-A](https://doi.org/10.1016/0896-6273(92)90239-A) [Medline](#)
45. M. Carroll, M. Levasseur, C. Wood, M. Whitaker, K. T. Jones, A. McDougall, Exploring the mechanism of action of the sperm-triggered calcium-wave pacemaker in ascidian zygotes. *J. Cell Sci.* **116**, 4997–5004 (2003). [doi:10.1242/jcs.00846](https://doi.org/10.1242/jcs.00846) [Medline](#)
46. S. Regot, J. J. Hughey, B. T. Bajar, S. Carrasco, M. W. Covert, High-sensitivity measurements of multiple kinase activities in live single cells. *Cell* **157**, 1724–1734 (2014). [doi:10.1016/j.cell.2014.04.039](https://doi.org/10.1016/j.cell.2014.04.039) [Medline](#)
47. A. Roure, U. Rothbacher, F. Robin, E. Kalmar, G. Ferone, C. Lamy, C. Missero, F. Mueller, P. Lemaire, A multicassette Gateway vector set for high throughput and comparative analyses in ciona and vertebrate embryos. *PLOS ONE* **2**, e916 (2007). [doi:10.1371/journal.pone.0000916](https://doi.org/10.1371/journal.pone.0000916) [Medline](#)
48. U. Krzic, S. Gunther, T. E. Saunders, S. J. Streichan, L. Hufnagel, Multiview light-sheet microscope for rapid in toto imaging. *Nat. Methods* **9**, 730–733 (2012). [doi:10.1038/nmeth.2064](https://doi.org/10.1038/nmeth.2064) [Medline](#)
49. N. Otsu, A threshold selection method from gray-level histograms. *Automatica* **11**, 23–27 (1975).
50. S. Ourselin, A. Roche, S. Prima, N. Ayache, “Block matching: A general framework to improve robustness of rigid registration of medical images” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*, S. L. Delp, A. M. DiGoia, B. Jaramaz, Eds., vol. 1935 of *Lecture Notes in Computer Science* (Springer, 2000), pp. 557–566.
51. L. Guignard, C. Godin, U.-M. Fiuza, L. Hufnagel, P. Lemaire, G. Malandain, Spatio-temporal registration of embryo images. *Proc. IEEE Int. Symp. Biomed. Imaging* **2014**, 778–781 (2014).
52. P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer, 1999).
53. D. Nicol, I. A. Meinertzhagen, Cell counts and maps in the larval central nervous system of the ascidian *Ciona intestinalis* (L.). *J. Comp. Neurol.* **309**, 415–429 (1991). [doi:10.1002/cne.903090402](https://doi.org/10.1002/cne.903090402) [Medline](#)
54. F. Amat, W. Lemon, D. P. Mossing, K. McDole, Y. Wan, K. Branson, E. W. Myers, P. J. Keller, Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nat. Methods* **11**, 951–958 (2014). [doi:10.1038/nmeth.3036](https://doi.org/10.1038/nmeth.3036) [Medline](#)
55. M. W. Kirschner, L. Shapiro, H. McAdams, G. Almouzni, P. A. Sharp, R. A. Young, U. Alon, Fifty years after Jacob and Monod: What are the unanswered questions in

- molecular biology? *Mol. Cell* **42**, 403–404 (2011). [doi:10.1016/j.molcel.2011.05.003](https://doi.org/10.1016/j.molcel.2011.05.003) [Medline](#)
56. P. Márquez-Neila, L. Baumela, L. Alvarez, A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 2–17 (2014). [doi:10.1109/TPAMI.2013.106](https://doi.org/10.1109/TPAMI.2013.106) [Medline](#)
57. G. Michelin, L. Guignard, U.-M. Fiuza, G. Malandain, Embryo Cell Membranes Reconstruction by Tensor Voting. *Proc. IEEE Int. Symp. Biomed. Imaging* **2014**, 1259–1262 (2014).
58. G. Medioni, M.-S. Lee, C.-K. Tang, *A Computational Framework for Segmentation and Grouping* (Elsevier Science, 2000).
59. E. G. Conklin, *The Organization and Cell-Lineage of the Ascidian Egg* (Academy of Natural Science, 1905).
60. A. G. Cole, I. A. Meinertzhagen, The central nervous system of the ascidian larva: Mitotic history of cells forming the neural tube in late embryonic *Ciona intestinalis*. *Dev. Biol.* **271**, 239–262 (2004). [doi:10.1016/j.ydbio.2004.04.001](https://doi.org/10.1016/j.ydbio.2004.04.001) [Medline](#)
61. M. Brozovic, C. Dantec, J. Dardaillon, D. Dauga, E. Faure, M. Gineste, A. Louis, M. Naville, K. R. Nitta, J. Piette, W. Reeves, C. Scornavacca, P. Simion, R. Vincentelli, M. Bellec, S. B. Aicha, M. Fagotto, M. Guérout-Bellone, M. Haeussler, E. Jacox, E. K. Lowe, M. Mendez, A. Roberge, A. Stolfi, R. Yokomori, C. T. Brown, C. Cambillau, L. Christiaen, F. Delsuc, E. Douzery, R. Dumollard, T. Kusakabe, K. Nakai, H. Nishida, Y. Satou, B. Swalla, M. Veeman, J.-N. Volff, P. Lemaire, ANISEED 2017: Extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res.* **46**, D718–D725 (2018). [doi:10.1093/nar/gkx1108](https://doi.org/10.1093/nar/gkx1108) [Medline](#)
62. J. Lindblad, Surface area estimation of digitized 3D objects using weighted local configurations. *Image Vis. Comput.* **23**, 111–122 (2005). [doi:10.1016/j.imavis.2004.06.012](https://doi.org/10.1016/j.imavis.2004.06.012)
63. D. Auber, D. Archambault, R. Bourqui, M. Delest, J. Dubois, A. Lambert, P. Mary, M. Mathiaut, G. Melançon, B. Pinaud, B. Renoust, J. Vallet, “TULIP 5” in *Encyclopedia of Social Network Analysis and Mining*, R. Alhajj, J. Rokne, Eds. (Springer, 2017), pp. 1–28.
64. P. Bille, A survey on tree edit distance and related problems. *Theor. Comput. Sci.* **337**, 217–239 (2005). [doi:10.1016/j.tcs.2004.12.030](https://doi.org/10.1016/j.tcs.2004.12.030)
65. K. Zhang, A constrained edit distance between unordered labeled trees. *Algorithmica* **15**, 205–222 (1996). [doi:10.1007/BF01975866](https://doi.org/10.1007/BF01975866)
66. P. Ferraro, C. Godin, A distance measure between plant architectures. *Ann. For. Sci.* **57**, 445–461 (2000). [doi:10.1051/forest:2000134](https://doi.org/10.1051/forest:2000134)
67. G. E. Noether, “A brief survey of nonparametric statistics,” *Studies in Statistics*, R. V. Hogg, Ed., vol. 19 of *Studies in Mathematics* (Mathematical Association of America, 1978), pp. 39–65.
68. O. Hertwig, Untersuchungen zur morphologie und physiologie der Zelle: das Problem der Befruchtung und der Isotropie des Eies, eine Theorie der Vererbung, vol. 3 (Fischer, 1884).
69. J. H. Ward Jr., Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963). [doi:10.1080/01621459.1963.10500845](https://doi.org/10.1080/01621459.1963.10500845)

70. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics Springer, 2009).
71. K. S. Imai, K. Hino, K. Yagi, N. Satoh, Y. Satou, Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: Towards a comprehensive understanding of gene networks. *Development* **131**, 4047–4058 (2004). [doi:10.1242/dev.01270](https://doi.org/10.1242/dev.01270) [Medline](#)
72. K. S. Imai, M. Levine, N. Satoh, Y. Satou, Regulatory blueprint for a chordate embryo. *Science* **312**, 1183–1187 (2006). [doi:10.1126/science.1123404](https://doi.org/10.1126/science.1123404) [Medline](#)
73. O. Tassy, D. Dauga, F. Daian, D. Sobral, F. Robin, P. Khoueiry, D. Salgado, V. Fox, D. Caillol, R. Schiappa, B. Laporte, A. Rios, G. Luxardi, T. Kusakabe, J.-S. Joly, S. Darras, L. Christiaen, M. Contensin, H. Auger, C. Lamy, C. Hudson, U. Rothbächer, M. J. Gilchrist, K. W. Makabe, K. Hotta, S. Fujiwara, N. Satoh, Y. Satou, P. Lemaire, The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. *Genome Res.* **20**, 1459–1468 (2010). [doi:10.1101/gr.108175.110](https://doi.org/10.1101/gr.108175.110) [Medline](#)
74. N. Haupaix, P. B. Abitua, C. Sirour, H. Yasuo, M. Levine, C. Hudson, Ephrin-mediated restriction of ERK1/2 activity delimits the number of pigment cells in the *Ciona* CNS. *Dev. Biol.* **394**, 170–180 (2014). [doi:10.1016/j.ydbio.2014.07.010](https://doi.org/10.1016/j.ydbio.2014.07.010) [Medline](#)
75. P. Dehal, Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H. G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B. I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh, D. S. Rokhsar, The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002). [doi:10.1126/science.1080049](https://doi.org/10.1126/science.1080049) [Medline](#)
76. C. Hudson, H. Yasuo, A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. *Development* **133**, 2855–2864 (2006). [doi:10.1242/dev.02466](https://doi.org/10.1242/dev.02466) [Medline](#)
77. S. Darras, H. Nishida, The BMP signaling pathway is required together with the FGF pathway for notochord induction in the ascidian embryo. *Development* **128**, 2629–2638 (2001). [Medline](#)
78. C. Hudson, S. Lotito, H. Yasuo, Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. *Development* **134**, 3527–3537 (2007). [doi:10.1242/dev.002352](https://doi.org/10.1242/dev.002352) [Medline](#)
79. K. Bartscherer, M. Boutros, Regulation of Wnt protein secretion and its role in gradient formation. *EMBO Rep.* **9**, 977–982 (2008). [doi:10.1038/embor.2008.167](https://doi.org/10.1038/embor.2008.167) [Medline](#)
80. T. Minokawa, K. Yagi, K. W. Makabe, H. Nishida, Binary specification of nerve cord and notochord cell fates in ascidian embryos. *Development* **128**, 2007–2017 (2001). [Medline](#)

81. V. Picco, C. Hudson, H. Yasuo, Ephrin-Eph signalling drives the asymmetric division of notochord/neural precursors in *Ciona* embryos. *Development* **134**, 1491–1497 (2007). [doi:10.1242/dev.003939](https://doi.org/10.1242/dev.003939) [Medline](#)
82. C. Alexandre, A. Baena-Lopez, J.-P. Vincent, Patterning and growth control by membrane-tethered Wingless. *Nature* **505**, 180–185 (2014). [doi:10.1038/nature12879](https://doi.org/10.1038/nature12879) [Medline](#)
83. C. M. Jones, N. Armes, J. C. Smith, Signalling by TGF- β family members: Short-range effects of Xnr-2 and BMP-4 contrast with the long-range effects of activin. *Curr. Biol.* **6**, 1468–1475 (1996). [doi:10.1016/S0960-9822\(96\)00751-8](https://doi.org/10.1016/S0960-9822(96)00751-8) [Medline](#)
84. S. Piccolo, Y. Sasai, B. Lu, E. M. De Robertis, Dorsoventral patterning in *Xenopus*: Inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589–598 (1996). [doi:10.1016/S0092-8674\(00\)80132-4](https://doi.org/10.1016/S0092-8674(00)80132-4) [Medline](#)
85. C. Hudson, P. Lemaire, Induction of anterior neural fates in the ascidian *Ciona intestinalis*. *Mech. Dev.* **100**, 189–203 (2001). [doi:10.1016/S0925-4773\(00\)00528-1](https://doi.org/10.1016/S0925-4773(00)00528-1) [Medline](#)
86. L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118**, 4495–4509 (2005). [doi:10.1242/jcs.02553](https://doi.org/10.1242/jcs.02553) [Medline](#)
87. K. Heinecke, A. Seher, W. Schmitz, T. D. Mueller, W. Sebald, J. Nickel, Receptor oligomerization and beyond: A case study in bone morphogenetic proteins. *BMC Biol.* **7**, 59 (2009). [doi:10.1186/1741-7007-7-59](https://doi.org/10.1186/1741-7007-7-59) [Medline](#)
88. S. Aykul, W. Ni, W. Mutatu, E. Martinez-Hackert, Human Cerberus prevents nodal-receptor binding, inhibits nodal signaling, and suppresses nodal-mediated phenotypes. *PLOS ONE* **10**, e0114954 (2015). [doi:10.1371/journal.pone.0114954](https://doi.org/10.1371/journal.pone.0114954) [Medline](#)
89. O. A. Ibrahimi, F. Zhang, S. C. Lang Hrstka, M. Mohammadi, R. J. Linhardt, Kinetic model for FGF, FGFR, and proteoglycan signal transduction complex assembly. *Biochemistry* **43**, 4724–4730 (2004). [doi:10.1021/bi0352320](https://doi.org/10.1021/bi0352320) [Medline](#)
90. J. P. Himanen, M. J. Chumley, M. Lackmann, C. Li, W. A. Barton, P. D. Jeffrey, C. Vearing, D. Geleick, D. A. Feldheim, A. W. Boyd, M. Henkemeyer, D. B. Nikolov, Repelling class discrimination: ephrin-A5 binds to and activates EphB2 receptor signaling. *Nat. Neurosci.* **7**, 501–509 (2004). [doi:10.1038/mn1237](https://doi.org/10.1038/mn1237) [Medline](#)
91. S. Dyson, J. B. Gurdon, The interpretation of position in a morphogen gradient as revealed by occupancy of activin receptors. *Cell* **93**, 557–568 (1998). [doi:10.1016/S0092-8674\(00\)81185-X](https://doi.org/10.1016/S0092-8674(00)81185-X) [Medline](#)
92. D. Sako, A. V. Grinberg, J. Liu, M. V. Davies, R. Castonguay, S. Maniatis, A. J. Andreucci, E. G. Pobre, K. N. Tomkinson, T. E. Monnell, J. A. Ucran, E. Martinez-Hackert, R. S. Pearsall, K. W. Underwood, J. Seehra, R. Kumar, Characterization of the ligand binding functionality of the extracellular domain of activin receptor type IIb. *J. Biol. Chem.* **285**, 21037–21048 (2010). [doi:10.1074/jbc.M110.114959](https://doi.org/10.1074/jbc.M110.114959) [Medline](#)
93. Z. Pei, N. E. Baker, Competition between Delta and the Abruptex domain of Notch. *BMC Dev. Biol.* **8**, 4 (2008). [doi:10.1186/1471-213X-8-4](https://doi.org/10.1186/1471-213X-8-4) [Medline](#)
94. J. Downward, The GRB2/Sem-5 adaptor protein. *FEBS Lett.* **338**, 113–117 (1994). [doi:10.1016/0014-5793\(94\)80346-3](https://doi.org/10.1016/0014-5793(94)80346-3) [Medline](#)
95. N. Haupaix, A. Stolfi, C. Sirour, V. Picco, M. Levine, L. Christiaen, H. Yasuo, p120RasGAP mediates ephrin/Eph-dependent attenuation of FGF/ERK signals during cell fate specification in ascidian embryos. *Development* **140**, 4347–4352 (2013). [doi:10.1242/dev.098756](https://doi.org/10.1242/dev.098756) [Medline](#)

96. N. Ohta, Y. Satou, Multiple signaling pathways coordinate to induce a threshold response in a chordate embryo. *PLOS Genet.* **9**, e1003818 (2013).
[doi:10.1371/journal.pgen.1003818](https://doi.org/10.1371/journal.pgen.1003818) [Medline](#)
97. U. S. Bhalla, P. T. Ram, R. Iyengar, MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* **297**, 1018–1023 (2002).
[doi:10.1126/science.1068873](https://doi.org/10.1126/science.1068873) [Medline](#)
98. W. Shi, M. Levine, Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the *Ciona* embryo. *Development* **135**, 931–940 (2008).
[doi:10.1242/dev.011940](https://doi.org/10.1242/dev.011940) [Medline](#)
99. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). [doi:10.1038/nmeth.2019](https://doi.org/10.1038/nmeth.2019) [Medline](#)
100. K. Kondoh, K. Kobayashi, H. Nishida, Suppression of macho-1-directed muscle fate by FGF and BMP is required for formation of posterior endoderm in ascidian embryos. *Development* **130**, 3205–3216 (2003). [doi:10.1242/dev.00521](https://doi.org/10.1242/dev.00521) [Medline](#)
101. H. Yasuo, C. Hudson, FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in *Ciona* embryos. *Dev. Biol.* **302**, 92–103 (2007).
[doi:10.1016/j.ydbio.2006.08.075](https://doi.org/10.1016/j.ydbio.2006.08.075) [Medline](#)
102. C. Hudson, H. Yasuo, Patterning across the ascidian neural plate by lateral Nodal signalling sources. *Development* **132**, 1199–1210 (2005). [doi:10.1242/dev.01688](https://doi.org/10.1242/dev.01688) [Medline](#)
103. K. S. Imai, N. Satoh, Y. Satou, An essential role of a FoxD gene in notochord induction in *Ciona* embryos. *Development* **129**, 3441–3453 (2002). [Medline](#)
104. K. S. Imai, N. Satoh, Y. Satou, Early embryonic expression of FGF4/6/9 gene and its role in the induction of mesenchyme and notochord in *Ciona savignyi* embryos. *Development* **129**, 1729–1738 (2002). [Medline](#)
105. E. Wagner, M. Levine, FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012). [doi:10.1242/dev.078485](https://doi.org/10.1242/dev.078485) [Medline](#)
106. H. L. Sladitschek, U.-M. Fiuza, D. Pavlinic, V. Benes, L. Hufnagel, P. A. Neveu, MorphoSeq: Full single-cell transcriptome dynamics up to gastrulation in a chordate. *Cell* **181**, 922–935.e21 (2020). [doi:10.1016/j.cell.2020.03.055](https://doi.org/10.1016/j.cell.2020.03.055) [Medline](#)