Supplementary Data

1.1 Introduction:

This model considers cells, cell parts, extra-cellular matrix (ECM), gene products and other molecules involved in gene regulation. The model includes in a unified framework bio-mechanical and gene product interactions in development. The model also includes all basic cell behaviors known in animal cells (Salazar-Ciudad et al, 2003). These are: cell growth (polar and non-polar), cell division (directed and non-directed), apoptosis, secretion of ECM and signaling molecules, reception of extracellular signals, cell contraction, cell adhesion and movement and shape change as a consequence of those.

The model includes mesenchymal cells, epithelial cells and ECM. Each cell can further differentiate to adopt a specific morphology and patterns of gene expression due to signaling and mechanical forces. Mesenchymal cells and ECM are made of spherical nodes while epithelial cells are made of cylinders. Each cylinder is made of two nodes, an apical and a basal one. The number, size and position (in a continuous 3D space) of nodes changes via model dynamics. As a result cells move and change their size and form. Changes in the spatial location and shape of cells configure the overall changes in an embryo's morphology.

All calculations, including molecule concentrations and diffusion, are made exclusively on the nodes but allowing diffusion of some molecules between nodes within and between cells (see below). The spatial distribution of those nodes represents the embryo's morphology and within each cell it represents cell shape. Each node has a set of mechanical properties and can accumulate different types of molecules. These properties are numerical values that affect the forces acting between nodes and are affected by the molecules present in nodes. Cells also have properties whose values are affected by the molecules present in each of its nodes.

In addition, the model includes some global parameters and a gene network that can be different depending on the organ or embryo part being studied. The gene network is specified by a n_g *x* n_g matrix (matrix *T*), a n_g *x* n_g *x* n_g matrix (matrix *R*), a n_g *x* n_{ga} matrix (matrix *E*) and a n_g *x* n_{gb} matrix (matrix *C*). n_g is the number of different molecules the model considers (a value specified by the user), n_{ga} is the number of node properties plus the number of cell properties and n_{gb} is the number of cell behaviors considered in the model. Each element, t_{ij}, in the *T* matrix specifies how strongly gene *j* activates (positive value) or inhibits (negative value) the transcription of gene *i*. Each element, r_{ijk}, in the *R* matrix specifies how strongly molecule *i* catalyzes the production of molecule *k* from molecule *j*. Thus, our model considers both transcription of genes and the modification of molecules that may be gene products (for example post-translational modifications of gene products) or not (such as many secondary messengers). We refer to both the gene products and other molecules that affect gene expression as *regulatory molecules*. Each element, e_{ij}, in the *E* matrix specifies how strongly molecule *i* increases or decreases node property *j* in the nodes where this molecule is present. Each element, c_{ij}, in the *C* matrix specifies how strongly gene *i* promotes or inhibits cell behavior *j* in the nodes where it is expressed. In addition each molecule has its own intrinsic rate of degradation and diffusion. Regulatory molecules can (i) diffuse between any of the nodes of a cell, (ii) only between the apical or basal nodes of an epithelial cell, or, (iii) in the case of growth factors, in the space between cells.

Cells also have properties whose values are affected by the regulatory molecules present in each of its nodes. Cell behaviors are implemented in the model as specific rules of manipulation of these nodes and their distribution between cells. Which behavior a cell performs in a given instant of time is affected by the regulatory molecules present in it (in a way specified by the user).

Our model does not impose any specific gene network but is a computational framework in which any arbitrary network can be implemented and in which these can regulate cell behaviors with realistic biomechanics. Thus, it is the user that decides which gene network and initial conditions the model uses according to the specific developmental system the user wants to study. The embryo editor and gene network software (http://www.biocenter.helsinki.fi/salazar/software.html) provide an easy way to do that without needing to program or understand the code (only the biological bases of the model need to be understood for that).

1.2 List of model elements:

1.2.1 Subcellular elements and nodes:

Subcellular elements are the smallest functional entities implemented in the model. They represent a physical portion of a cell. Mesenchymal subcellular elements are implemented as spherical elastic volumes, as has been done in the Subcellular Elements Model or SEM (Newman 2005), and from now on they will be called mesenchymal nodes. In contrast, epithelial cells are made of sub-cellular elements with cylindrical shape. Each cylinder represents a portion of the cell extending from the apical surface to the basal surface. Thus, there are 3 different surfaces in each cylinder: apical and basal surfaces, which are in contact with the extracellular space or with mesenchymal nodes, and a lateral surface, which contacts the lateral surface of other cylinders from the same cell or the ones from other cells (Supplementary Fig. 3B). Each epithelial cylinder consists of two nodes, an apical and a basal one, which may have different properties. These two nodes are tied by an elastic spring (see Section 2.2.1). In order to represent the ECM a third type of subcellular element is implemented as spherical nodes that do not belong to any cell (see Section 2.3).

Node properties are named by a lower case *p* and a three letter superindex specific of each property and a second subindex for the specific node (e.g: p_i^{EQD} is the property "equilibirum distance" for node *i*; notice the superindex is not p_i multiplied *EQD* times).

1.2.2 Cells.

Cells are functional entities in the model that can change their shape and perform a wide range of cell behaviors (see Section 5). Cell properties are named by a upper case *P* and a three letter superindex specific of each property and a second subindex for the specific cell. Thus, for example, P_i^{PHA} is property "PHA" of cell *i*.

Each cell is composed of one or several subcellular elements and the shape of the cell, thus, is given by the relative positions of those elements. The number of nodes in a cell in a given moment depends on how many nodes it had initially, on how much has it grown, on whether it has divided or not and on two cell properties. The first, P^{MIN}, specifies the minimal number of nodes in a cell, this precludes cell divisions leading to any daughter cell with less than that number. The other one, P^{MAX}, specifies the maximal number of nodes a cell can have (beyond that number the cell may divide, if it is in right cell cycle stage, or stop growing). In those developmental systems where cell shape is not important the user can choose to have mesenchymal cells made of a maximum of only one node and epithelial cells of two nodes, a cylinder (this requires, in addition, setting the logical model parameter, L_1, to 1). These cell properties, like all other properties, can be genetically regulated. Each cell has one node specified as the nucleus, which only differs from the other nodes in that it's the only node in which transcription takes place. The fact that cells are made of nodes permits cells to have internal spatial asymmetries, that is, different nodes in a cell can have different properties (either different mechanical properties or different amounts of different regulatory molecules).

1.2.3 Gene products, regulatory molecules and genetic parameters.

The amount of a regulatory molecule at a node can change during simulation time due to gene product transcription, diffusion between nodes, biochemical reactions and degradation (see Section 3). Each regulatory molecule has a set of properties, which we call generically gene parameters even if not all the regulatory molecules are necessarily gene products. These properties are actual model parameters in the sense that they are assumed to be genetically encoded and, thus, do not change during the simulation. Chemical transformation of regulatory molecules is implemented in the model as the transition of one type of molecule to another one by regulated catalysis. Thus, transformed molecules are not considered to have changed properties but abundance (this abundance of each molecular species being a property of each node). The only difference the model considers between gene products and other regulatory molecules is that gene products can be transcribed and translated while other regulatory molecules can only be transformed (we include different equations for transcription and catalysis of each chemical transformation). Transcription and translation can be considered together by the transcription equations or can be considered separately by using the non-transcription catalysis equations for translation. Gene products can also be modified post-translationally by reaction catalysed by other regulatory molecules. In this model we treat each one of these modifications as a different gene product that can have its own genetic parameters. A specific gene product modification may be present or not in a node depending on the model dynamics and on the initial conditions.

These molecular parameters include diffusivity, rate of degradation, the ability to interact with other molecules (chosen by the user according to the system being studied) as specified in the *T* and *R* matrices and their regulation of node properties (*E* matrix) and cell behaviors (*C* matrix). In addition, the model includes a *B* matrix that specifies the affinity of membrane molecules that mediate cell-cell adhesion. By choosing the values of those parameters, for example by using the NetworkMaker software, the user can implement the number of genes and the the gene network he/she wants to study and how that affects node properties and cell properties and behaviors.

1.2.4 *Global model parameters:* These are numerical values which, like the molecular parameters, do not change during a model simulation but that can be set to different values in different simulations of the model and do not have a direct correspondence with anything genetic. These include things such as the temperature and *logic model parameters* that specify some details about how the model is actually numerically implemented. These logic model parameters are indentified by an *L* with a subindex specific to each parameter. Other model parameters are represented by an *M* with a subindex specific to each parameter (see section 9.2).

1.2.5 *Initial conditions*: Those are the numerical values of all the nodes, and all the node and cell properties at time zero of a simulation (thus includes its location in 3D space and the amount of each regulatory molecule present in each node). These can be changed from simulation to simulation (with the embryo editor software or by manually editing the files that contain them). They are simply the stage in development from which we want to start to simulate development. This can be from a single cell with some internal spatial asymmetries (e.g., a zygote), or any arbitrary later stage in development (for example the second instar wing disc of *Drosophila*). Each initial condition is thus what we call a developmental pattern (Salazar-Ciudad *et al.* 2003) and the model dynamics transform the initial pattern into a different later pattern according to the genetic parameters and model parameters. These parameters (e.g., the genes and gene networks that regulate which cell behaviours) and the initial conditions are subject to change between simulations depending on the question and the developmental system being studied by the user.

1.3. Mechanical Forces

1.3.1 Node neighboring and basic biomechanical interactions:

Nodes have a size, specified by a radius in both spheres and cylinders. Two nodes adhere to each other if they come into contact (Supplementary Fig. 3A), that is, if their distance is smaller than the sum of their radii. The radius is a node property, p^{ADD}. Adhesion brings these nodes closer, further decreasing their distance until the equilibrium distance between nodes is reached. This represents the generic property of cell adhesion. If the nodes are from different cells this adhesion can be increased or decreased by adhesion molecules expressed at the nodes, as we will later explain.

Cell parts such as nodes represent physical objects and thus two nodes cannot occupy the same spatial location. Nodes, thus, have a second radius, node property p^{EQD}, and two nodes start repelling each other if the distance between them is shorter that the sum of their p^{EQD} (Supplementary Fig. 3A). If the distance between two nodes is exactly equal to the sum of their p^{EQD} radii then an equilibrium is reached in which these nodes neither attract nor repel each other.

From these forces the movement of node *i* due to its interaction with other nodes is:

$$\frac{\partial \vec{r}_i}{\partial t} = \sum_{j=1}^{j=n_d} f_{Aij} \ \hat{u}_{ij} \quad (1)$$

Where n_d is the number of nodes in the embryo, $\vec{r}_i$ is the position in three-dimensional space of node *i*, *t* is time (the model uses continuous time), f_{Aij} is the force modulus and $\hat{u}_{ij}$ is the unit vector between node *i* and node *j* for spherical nodes and an analogous property for cylinders. We assume that most developmental processes happen within highly viscous media (Purcell 1977), thus we calculate movement through an over-damped equation of motion.

The force modulus between node *i* and *j* is:

$$\begin{cases} f_{Aij} = k_{ij}^{REP}\left(d_{ij} - d_{ij}^{EQD}\right) & if \quad d_{ij} < d_{ij}^{EQD} \\ f_{Aij} = k_{ij}^{YOU}\left(d_{ij} - d_{ij}^{EQD}\right) & if \quad d_{ij} \le d_{ij}^{EQD} \le d_{ij}^{ADD} \\ f_{Aij} = 0 & if \quad d_{ij} > d^{ADD} \end{cases} \quad (2)$$

Where d_{ij} is the the distance at which node *i* and *j* are, d_{ij}^{EQD} is the equilibrium distance between node *i* and node j (simply the sum of the equilibrum radii, *EQD*, node properties of nodes *i* and *j*, Supplementary Fig. 3A) and d_{ij}^{ADD} is the sum of the node property *ADD* of node *i* and *j*:

$$\begin{aligned} k_{ij}^{REP} &= p_i^{REP} + p_j^{REP} \\ k_{ij}^{YOU} &= p_i^{YOU} + p_j^{YOU} \\ d_{ij}^{EQD} &= p_i^{EQD} + p_j^{EQD} \\ d_{ij}^{ADD} &= p_i^{ADD} + p_j^{ADD} \end{aligned} \quad (3)$$

p^{REP} and p^{YOU} are bio-mechanical properties of the nodes that specify how strong per unit distance are the repulsion and elasticity forces respectively between a pair of nodes. If *i* and *j* belong to different cells, k^{ADH} is used instead of k^{YOU} and k^{REC} is used instead of k^{REP}. That is, we implement intercellular adhesion as an elastic force between cells. p^{REC} is different from p^{REP} because naturally cells may more strongly resist incoming matter from other cells than from the same cell. Between cells, in addition to the generic adhesion between nodes, there is an adhesion term that depends on which adhesion molecules are expressed in each of the nodes.

$$k_{ij}^{REC} = p_i^{REC} + p_j^{REC}$$
$$k_{ij}^{YOU} = p_i^{ADH} + p_j^{ADH} + \sum_{l=1}^{n_{ADH}} \sum_{q=l}^{n_{ADH}} \left(g_{il} g_{jq} B_{lq} \right)$$

(4)

Where, n_{ADH} is the number of different types of adhesion molecules (a model parameter), g_{il} is the amount of adhesion molecule l in node i, g_{iq} is the amount of adhesion molecule q in node j and B_{lq} is the affinity coefficient between adhesion molecules l and q (genetic parameter). Notice that homotypic adhesion is allowed between different nodes. B_{pq} can be negative causing cells which express the p and q adhesion molecules to repel each other (as happens with some semaphorins (Bagnard et al. 1998), and ephrins (Wang & Anderson 1997). Thus, the adhesion force is simply the product of how many adhesion molecules there are of each binding pair of a giventype and the affinity of each pair of types of adhesion molecules.

If node density in space and the p_i^{ADD} of nodes are both large it becomes possible that two nodes could interact even if there are nodes between them. To avoid that unrealistic situation the model allows for three alternative algorithms to determine which nodes can effectively interact. Each simulation should be run with only one of these alternative methods:

1. Simple method: As described above, any node j that is at a distance (in 3D space) smaller than $p_i^{ADD} + p_j^{ADD}$ from a given node (node i) is interacting with it.

2. Delaunay method: A tesselation of the 3D space is performed by the Delaunay triangulation algorithm (as in Delile et al. 2013), taking each node as a vertex. Then only the nodes that are connected by an edge in this tesselation and are at a distance smaller than the sum of their p^{ADD}, as above, interact.

3. Gabriel graph based method: Similar to method 1, but for every two nodes at the right distance an additional criterion needs to be fulfilled. A sphere of a diameter equal to the distance between the two nodes is put in the mid-line between them. If there is any node within this sphere then those two nodes do not interact.

Method 1 is computationally faster and is the one used in the SEM (Newman 2005). Method 3 is used in some similar models (Delile et al. 2013). In most situations these three methods provide very similar results. Only when nodes extensively overlap in space (because overall large values in the p^{ADD} property in respect to node density in space) it is more realistic to use method 2 or 3 (method 3 being computationally more advantageous for only a slightly lower degree of realism). These two last methods simply preclude the interaction between two nodes that are close enough if there are other nodes in between. In that sense they implement a screening between nodes. These alternatives are specified, respectively by the logic model parameters L_3 and L_{11}.

1.3.2 Forces in epithelial nodes:

In epithelia, the above forces work in a different way (although the rules to define neighborhood, which nodes interact with which, is the same). Equations (1) and (2) are the ones used to calculate the forces in cylinders but the $\hat{u}_{ij}$, the direction of the force, and d_{ij} the distance between elements is calculated differently to take into account that epithelial elements have a cylindrical shape.

1.3.2.1 Direction of the mechanical interactions between epithelial cylinders.

Epithelial cylinders have two types of surface: apical/basal and lateral. Those surfaces are defined by the orientation of the cylinder vector ($\vec{s}_{ik}$) connecting the two epithelial nodes (i and k) composing a cylinder.

$$\vec{s}_{ik} = \vec{r}_i - \vec{r}_k \quad (5)$$

Where r_i and r_k are the position vectors of nodes i and k respectively. Apical/basal surfaces can interact with mesenchymal or ECM nodes or with the apical/basal surface of other epithelial cylinders. Lateral surfaces can only interact with lateral surfaces of other cylinders. (Supplementary Fig. 3B,C,D).

Case 1: Apical/basal cylinder surface against spheric node. (Supplementary Fig. 3C)

$$\vec{u}_{ij} = -\vec{s}_{ik} \quad (6)$$

In this case equation (1) applies but $\vec{u}_{ij}$ is calculated with equation (6). This simply reflects the fact that the contact area between the apical/basal side of the cylinder and the spheric node is always parallel to the apical/basal surface of the cylinder (FIG1.C).

Case 2: Apical/basal surface against apical/basal surface of different cylinders. (Supplementary Fig. 3D)

In this case the interaction is calculated in a more complicated way. Let $\vec{m}_{ijkl}$ be a vector contained within the plane defined by the contact surface between the two cylinders (Supplementary Fig. 3D), $\vec{m}_{ijkl}$ is calculated as the sum of the spring vectors from the two cylinders ($\vec{s}_{ik}$ and $\vec{s}_{jl}$),

$$\vec{m}_{ijkl} = \vec{s}_{ik} + \vec{s}_{jl} \quad (7)$$

Where *i* and *k* are the nodes composing the first cylinder and *j* and *l* the nodes composing the second cylinder. $\vec{u}_{ij}$ has to be normal to the contact surface plane and at the same time has to be contained within the plane defined by $\vec{m}_{ijkl}$ and $\vec{c}_{ij}$ (Supplementary Fig. 3D),

$$\vec{u}_{ij} = \vec{c}_{ij} - \vec{m}_{ijkl} \frac{\vec{m}_{ijkl} \cdot \vec{c}_{ij}}{|\vec{m}_{ijkl}|^2} \quad (8)$$

Where $\vec{c}_{ij}$ is the vector connecting the two interacting nodes *i* and *j* (that is between nodes in the two different cylinders) (Fig. 1D). || denotes modulus. The $\vec{u}_{ij}$ calculated in equation (8) is then divided by its modulus (to become a unit vector) and then feed into equation (1) to calculate the force as before.

There is a special case, when s_{ik} and s_{jl} are parallel and opposite in orientation, then $\vec{m}_{ijkl}$ is equal to 0. In that case, since the contact surface is normal to both $\vec{s}_{ik}$ and $\vec{s}_{jl}$, equation (6) is used.

Case 3: Cylinder lateral surface against cylinder lateral surface (Supplementary Fig. 3B).

This case is similar to case 2, since $\vec{m}_{ijkl}$ (the sum of the two cylinder vectors) is contained within the contact surface plane and $\vec{c}_{ij}$ has to be orthogonal to $\vec{m}_{ijkl}$ and contained in the plane defined by $\vec{m}_{ijkl}$ and $\vec{c}_{ij}$. Thus in this case $\vec{u}_{ij}$ follows equation (8).

Note that, following Newton's third Law, the force acting on *j* must have the same modulus but opposite sign. Thus, in all cases,

$$\vec{u}_{ij} = -\vec{u}_{ji}$$

1.3.2.2 Distance between cylinders and between cylinders and spheres.

Since the vector $\vec{u}_{ij}$ used to calculate mechanical forces is not always equal to $\vec{c}_{ij}$, the distance measure, d_{ij}, used to calculate the force modulus cannot be the actual distance between nodes *i* and *j*, which is the length of $\vec{c}_{ij}$. That is why the effective distance d_{ij} in equation (2) is calculated in the case of cylinder as the projection of $\vec{c}_{ij}$ over $\vec{u}_{ij}$ (Supplementary Fig. 3).

$$d_{ij} = \frac{\vec{c}_i \cdot \vec{u}_{ij}}{|\vec{u}_{ij}|} \quad (10)$$

This simply reflects that in a cylinder the distances between interacting nodes are different (due to the cylindrical shape, Supplementary Fig. 3).

1.3.2.3 Additional forces in epithelial cylinders.

In addition epithelial nodes have two additional forces that are not present in other types of nodes:

1.3.2.3.1. *Spring (Supplementary Fig. 4A)*: The spring exerts an elastic force on the two nodes depending on the distance between them,

$$\vec{f}_{Sij} = k_{ij}^{HOO}\left(d_{ij} - p_i^{EQS}\right)\hat{s}_{ik} \quad (11)$$

Where $k_{ij}^{HOO}=p_i^{HOO} + p_j^{HOO}$ is the elastic coefficient of the spring (note p^{HOO} is a node property), p_i^{EQS} is the equilibrium length of the spring between node *i* and *j* (note $p_i^{EQS} = p_j^{EQS}$) and $\hat{s}_{ij}$ is the cylinder vector, the unit vector in the direction between both nodes in a cylinder. d_{ij} is simply the distance between node *i* and *j*.

1.3.2.3.2 *Epithelial bending forces (Supplementary Fig. 4B,C)*: We define two different forces that deal with bending of epithelial sheets. Due to external forces (e.g. a punctual force pushing one cylinder from the basal side) the apical or basal side of a given cylinders may slide in the apical-basal direction respect neighboring cylinders (Supplementary Fig. 4B,C). If we consider the relative position of the apical/basal nodes as a discrete representation of the continuous apical/basal epithelial surface, any displacement in the apical-basal direction will imply a local curvature of the epithelium. Thus, a radial force has to be defined that reduce local curvature. We define c_{ij} as the vector connecting neighboring node *i* and *j*, s_{ik} and s_{jl} as the vectors that define the elastic link to their basal counterparts and m_{ijkl} as the sum of s_{ik} and s_{jl} which defines the vector normal to the apical or basal surface between *i* and *j*. The force vector f_{ESTij} is calculated as,

$$\begin{cases} \vec{f}_{ESTij} = k_{ij}^{EST}\dfrac{\vec{m}_{ijkl}\cdot\vec{c}_{ij}}{|\vec{m}_{ijkl}|}\hat{m}_{ijkl} & if \ \dfrac{\vec{m}_{ijkl}\cdot\vec{c}_{ij}}{|\vec{m}_{ijkl}|} \geq M_{AMX}|\vec{c}_{ij}| \\ \vec{f}_{ESTij} = 0 & if \ \dfrac{\vec{m}_{ijkl}\cdot\vec{c}_{ij}}{|\vec{m}_{ijkl}|} < M_{AMX}|\vec{c}_{ij}| \end{cases} \quad (12)$$

This force always acts on the direction of the $\vec{m}_{ijkl}$ vector, and is proportional to the deviation of the angle formed by $\vec{m}_{ijkl}$ and $\vec{c}_{ij}$ from 90º (the angle found when to cylinders are totally aligned). k_{ij}^{EST} is derived from a specific node property ($k_{ij}^{EST}=p_i^{EST}+ p_j^{EST}$).

In order to minimize the sensitivity to small perturbations, we set a minimal value of the projection of $\vec{c}_{ij}$ over $\vec{m}_{ijkl}$ to apply $\vec{f}_{EST\ ij}$, that is when the projection of $\vec{c}_{ij}$ over $\vec{m}_{ijkl}$ is less than the product of the distance between the two nodes and M_{AMX}, a model parameter, no surface tension force is applied.

If the force generating the local curvature is persistent, the epithelial surface will remain curved, and thus the epithelial cylinders will have to reorient in order to make their apical-basal vector s normal to the local epithelial surface, which is approximated by the vectors connecting the apical/basal node with their apical/basal neighbors. This is accomplished by applying a rotational force that will rotate the cylinders until their apical-basal vector s is normal to the surface plane at that position. The force vector f_{ERPij} is calculated as,

$$\begin{cases} \vec{f}_{ERPij} = p_i^{ERP}\dfrac{\vec{s}_{ik}\cdot\vec{c}_{ij}}{|\vec{s}_{ik}|}\hat{c}_{ij} & if \ \dfrac{\vec{s}_{ik}\cdot\vec{c}_{ij}}{|\vec{s}_{ik}|} \geq M_{AMX}|\vec{c}_{ij}| \\ \vec{f}_{ERPij} = 0 & if \ \dfrac{\vec{s}_{ik}\cdot\vec{c}_{ij}}{|\vec{s}_{ik}|} < M_{AMX}|\vec{c}_{ij}| \end{cases} \quad (13)$$

This force is proportional to the deviation of the angle formed by $\vec{s}_{ik}$ and $\vec{c}_{ij}$ from 90º, but in this case the direction of the force is parallel to $\vec{c}_{ij}$, thus promoting a tilting of the epithelial cylinder that reaches an equilibrium (that is the force modulus becomes 0) when the apical-basal axis of the epithelial cylinder is normal to the apical/basal cell surface. k_{ij}^{ERP} is derived from a specific node property ($k_{ij}^{ERP}=p_i^{ERP}+ p_j^{ERP}$).

In order to minimize the sensitivity to small perturbations, we set a minimal value of the projection of $\vec{c}_{ij}$ over $\vec{s}_{ik}$ to apply $\vec{f}_{ERPij}$, that is when the projection of $\vec{c}_{ij}$ over $\vec{s}_{ik}$ is less than the product of the distance between the two nodes and M_{AMX}, a model parameter, no rotational force is applied.

Note that these forces act equally between any cylinders irrespective of whether these cylinders belong to the same cell or not. In that sense this force is a property of the epithelium as such.

In summary, thus, the forces acting on an epithelial node are:

$$\frac{\partial \vec{r}_i}{\partial t} = \vec{f}_{Sik} + \sum_{j=1}^{n_d} \left(f_{Aij} \ \hat{u}_{ij} + \vec{f}_{ESTij} + \vec{f}_{ERPij} \right) \quad (14)$$

where *k* is the node in the same cylinder than i and the sum is made over all the neighboring nodes except for *k*.

This initial arrangement and the forces just described allow these cells to mechanically behave as epithelia for a broad range of realistic model parameters and initial conditions. Thus, epithelia behave as in two and a half dimensions, they can fold in complex ways in three-dimensional space but they retain a basically two dimensional structure with all cells in an epithelium binding to each other in their lateral sides (the apical and basal sides can bind to non-epithelial nodes or to nodes from different epithelium or even to a fold of the same epithelium).

1.3.3. Extracellular matrix:

The ECM is represented by spherical nodes that do not belong to a cell. They interact with other nodes in the same fashion as mesenchymal nodes. ECM is usually composed of large proteoglycans and glycoproteins that swell the extracellular fluids and behave as a gel. Therefore by implementing the ECM as free nodes we capture their most relevant mechanical properties. ECM nodes can be secreted by cells (See below), they can also contain regulatory molecules and can harbor catalysis (but not transcription).

1.3.4. Node movement and noise

In addition to the movement equation defined at equation (14) there is some noise in node movements. At each time step, a proportion M_{NOI} (a model parameter) of the nodes are chosen at random and are tentatively moved in a random direction for a random distance between 0 and p_i^{DMO} , a mechanical property of each node. For each node the potential mechanical energy is calculated, by integrating the same force equations shown in section 2.1 and 2.2, in the new position. If the potential energy in the new position is smaller than in the old position the movement is accepted. If not, the movement is accepted with a probability proportional to the difference in potential energy between the new and old positions and inversely proportionally to a temperature parameter, model parameter M_{TEM}, plus a node property defining the node's propensity to movement (p^{MOV}),

$$\begin{cases} P_{accept} = e^{-\frac{U_{after} - U_{before}}{M_{TEM} + P_i^{MOV}}} & if \ \ U_{after} - U_{before} > 0 \\ P_{accept} = 1 & if \ \ U_{after} - U_{before} \leq 0 \end{cases} \quad (15)$$

where P_{accept} is the probability of realization of the movement, U_{before} is the potential energy in the node position before movement and U_{after} is the potential energy after the movement. If the movement is not accepted the node is put back to its old position. This energy biased noise reflects the fact that noise can affect nodes' positions but it is unlikely to bring nodes into very energetically unfavorable positions (e.g. noise is very unlikely to bring a node from a cell inside another cell). This is a standard way to implement noise in many physical and biological systems (such as in SEM and in the Pott's model (Graner & Glazier 1992).

At the level of cells and nodes this noise property, p^{MOV}, reflects in part the tendency of cells, especially mesenchymal cells, to temporarily extent and retract cytoplasmatic projections (filopodia, pseudopodia and related structures) into the extracellular space. The likelihood of a pseudopodium retracting after being extended depends on whether it finds a suitable strong binding site (either in other cells or in the substratum). Also different types of cells tend to have pseudopodia of different lengths and tend to extend them with different frequencies. In individually migrating cells, in addition, the binding of those extensions is also relatively unstable so that cells can dynamically move over space. In our model this is captured by the p^{MOV} and p^{DMO} node properties . The movement of a node by noise

can be represented as this node being the tip of a pseudopodium. p^{DMO} specifies how long pseudopodia can extend before being retracted and p^{MOV} specifies how labile this node binding is (the effect is simply to add to noise in eq 15). Each node would then bind according to its p^{ADH}, plus the amount adhesion molecules expressed both in that node and in the one it is making contact with.

Noise allows the migration of individual cells over space (this is not possible in the model without noise). The two corresponding node properties are affected, as are all other node properties, by the regulatory molecules contained in nodes. This allows the implementation of individual cell chemotaxis. Basically, extracellular signal gradients induce intracellular regulatory gradients that differentially affect various properties within the cell. Similarly, haptotaxis is implemented given a spatial gradient adhesion molecules since motile cells will migrate towards regions where the concentration of adhesion molecules is higher (See Fig. 5).

1.4. Gene expression, regulatory molecules and gene networks

Each regulatory molecule has a set of properties associated with it (which we call genetic parameters even if not all the molecules considered are gene products). These include the diffusivity of the molecule (D_i) and its intrinsic degradation rate (μ_i) and how they affect transcription, catalysis, node properties and cell behaviors. Each molecule, thus, has one row of its own in the T, R, E and C matrices.

The T, R, E and C matrices are the quantitative specification of the developmental mechanism used in a simulation (and thus would be different depending on the developmental system being simulated with the model).

In the model molecules can affect the concentration of each other in a node in two ways. Some regulatory molecules can affect the production *de novo* of gene products by affecting the transcription of genes and the translation of transcripts (these two processes can be considered together or separately with translation being a different kind of reaction). Some regulatory molecules can catalyze the reactions by which a certain type of molecule is transformed into another type of molecule. These latter reactions can be seen either as post-translational modifications of proteins, reactions involving non-protein molecules or simple binding of molecules leading to conformational changes in those (that then can change the properties of those molecules). Each different molecule can have different genetic parameters (e.g.: they can affect other regulatory molecules in different ways or regulate different node properties or cell behaviors). Thus, for example, in the model a certain protein and the same protein in a phosphorylated state are two distinct regulatory molecules that can have different genetic and regulatory properties (specified in different rows in the T,R,E and C). The same can happen when a regulatory molecule is bound to some other molecules. Binding can lead to conformational changes that may drive changes on the interactions of that regulatory molecule with other molecule and thus the bound form is considered to be a different molecule than the unbound form and then can have different genetic parameters (such as in T, R, E and C). This is often the case for receptors, the ligand unbound and bound forms of a receptor often can interact with different molecules.

1.4.1 Types of regulatory molecules.

In our model the function of a regulatory molecule is specified by their genetic parameter matrices. Thus transcription factors are gene products that affect transcription of other genes. This implies that they have non-zero elements in their row in the T matrix specifying which gene products regulate other genes' transcription.

Enzymes are regulatory molecules that have at least one non-zero element in their row in the R matrix. These can represent regulatory molecules involved in signal transduction such as kinases, proteases, lipases, etc... or metabolic enzymes. The R specifies which regulatory molecules catalyze the reaction between which regulatory molecules and with which activity per molecule. Regulatory molecules that catalyze the binding between molecules can also have non-zero elements in their row in the R matrix.

Regulatory molecules with a non-zero element in their row in the E matrix are regulating node properties. They can be for example myosins affecting contraction and thus p_i^{EQD} in specific nodes or molecules binding to the cytoskeleton to make nodes more stiff (p_i^{YOU}), etc...

In the same way regulatory molecules with non-zero elements in the C matrix are involved in regulating cell behaviors. These can be activators of mitosis, apoptosis, matrix secretion, etc...

Membrane adhesion molecules have a non-zero element in the first column of the E matrix. This element indicates the index of the adhesion molecule this regulatory molecule is. Then the B matrix indicates the affinity of binding between each adhesion molecule (each element in the B matrix is thus a model parameter as is the case for the other matrices).

There is nothing in the model restricting a regulatory molecule from being a transcription factor and an adhesion molecule at the same time or an enzyme and a transcription factor. All these functions are simply specified by the values in the T, R, E and C matrices. In addition, there is a genetic property called type. Each regulatory molecule can be of only one type. These are:

1.4.1.1 Extracellular diffusible signals or growth factors:

These are regulatory molecules that can diffuse between nodes in different cells or between ECM nodes, as well as between adjacent nodes belonging to the same cell in contact with the cell surface. The amount of an extracellular signal in a node in the model does not represent how much of it there is the node but how much of it there is the extra-cellular space around the node. They cannot diffuse across the apical basal axis of epithelial cells, that is between an apical node and a basal node of the same cylinder. These diffusible signals affect expression in a cell only be binding to specific receptors membrane receptors. The bound forms may have different genetic parameters than the unbound form (for example it may have a different catalytic specificity or directly affect some node property). Note that mesenchymal cells tend to have rather irregular, fibroblast-like, shapes and thus most nodes should be expected to have some contact with the extra-cellular space. In the case of epithelial cells all nodes are in contact with the extra-cellular space, but extracellular signals cannot diffuse between nodes within the same cylinder.

1.4.1.2 Ligand binding receptors:

They can only diffuse between adjacent nodes of the same cell in contact with the cell surface. Thus they cannot diffuse across the apical-basal axis of an epithelial cell. They can bind to extracellular diffusible signals located in the same node (if they are specified to do so) and form a receptor-ligand complex. The receptor-ligand complex may have different genetic parameters than the free receptor form.

1.4.1.3 Membrane tethered regulatory molecules:

They can only diffuse between adjacent nodes of the same cell in contact with the cell surface. Thus they cannot diffuse across the apical-basal axis of an epithelial cell. They can mediate signaling by cell contact, thus they can interact and bind to membrane bound receptors located in adjacent nodes belonging to different cells.

1.4.1.4 Apically or basally localized molecules.

These are intracellular regulatory molecules that are continuously being transported from one side of an epithelial cell (apical or basal) to the other. Molecules being transported to the apical side can freely diffuse between adjacent nodes from the same side, but are only transported along the the apical-basal axis of the cylinder when they are in the basal side. Molecules being transported to the basal side can diffuse freely between adjacent nodes from the same side, but are only transported along the apical-basal axis of the cylinder if they are in the apical side.

1.4.1.5 All other molecules

All other molecules can diffuse between all the nodes in a cell, but not between nodes of different cells.

1.4.2 Transcription (Fig. 2a).

Transcription can only happen in the cell nucleus, which is located in a specific node within the cell. The rate of transcription of gene *k* in node *i* (provided that *i* is a nuclear node) is:

$$Q_{ik} = \frac{\Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)}{1+\Phi\left(\sum_{l=1}^{n_g} t_{lk} g_{il}\right)} \quad (16)$$

Where Q_{ik} is the rate of transcription of gene *k* in node i, g_{il} is the amount of transcriptional factor *l* in node *i* and each t_{lk} term is the strength by which each specific transcriptional factor *k* activates (positive t_{lk}) or inhibits (negative t_{lk}) the transcription of gene *l* (each of them is an element of matrix *T*). The sum is done through all the regulatory molecules and by definition only transcriptional factors have t_{lk} terms different from zero. Φ is a function that is equal to 0 for values of x smaller than 0 and equals to x when x is greater than 0 ($\Phi(x)=0$ if $x<0$ and $\Phi(x)=x$ if $x>0$). This function is used to ensure that there is not such a thing as negative transcription (although t_{lk} can be negative and thus repress transcription).

The equation (16) represents the binding of several transcriptional factors to the promoter of gene *k*. This is a saturating process that, for simplicity, is represented by a Hill equation of order 1. This means that when there are few activator factors the rate of transcription increases with the amount of these factors. But when there are many of these factors the rate of transcription does not increase as much with the amount of activator factors since the binding sites for each of them in the promoter are likely to be already occupied by them. The same (Salazar-Ciudad et al. 2000, 2001) or similar (Reinitz & Sharp 1995) equation has been used in previous models of gene networks in development.

1.4.3 Non-transcriptional catalysis (Fig. 2b).

The rate of production of a regulatory molecule in a node depends on the product of the amount of each regulatory molecule that gives rise to it by the amount of each regulatory molecule that promotes this catalysis (all within the same node) (Fig. 2). The rate of production of a regulatory molecule *k* in node *i* is thus:

$$S_{ik} = W_{ik} - U_{ik} = \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jlk}\ g_{ij}\ \frac{g_{il}}{1+g_{il}} - \sum_{l=1}^{n_g}\sum_{j=1}^{n_g} r_{jkl}\ g_{ij}\ \frac{g_{ik}}{1+g_{ik}} \quad (17)$$

The first term W_{ik} defines the rate of production regulatory molecule *k* in node *i* due to the transformation of other forms *l* into *k* catalyzed by *j*. The second term U_{ik} defines the rate of loss of form *k* due to its transformation into other forms *l* mediated by catalyzation from *j*. The R matrix element r_{jlk} specifies the catalytic activity of regulatory molecule *j* on the transformation of regulatory molecule *l* into regulatory molecule *k*. Each term follows Michaelis-Menten kinetics in which K_M equals 1.

1.4.4 Receptor-ligand binding (Fig. 2c).

The kinetics of receptor-ligand binding need to be implemented differently, since in this case the two reactants (receptor and ligand) give rise rise to a single product (the receptor-ligand complex). Thus, the receptor-ligand complex is represented by a single regulatory molecule *c* which can be formed by binding the ligand, *l*, and the receptor, *k*, and can also be dissociated giving rise to *l* and *k*. The rate of change of *c* on node *i* is,

$$\begin{cases} S_{ic} = a_1 g_{il} g_{ik} - a_{-1} g_{ic} \\ a_1 = r_{clc} = r_{ckc} \\ a_{-1} = r_{ccl} = r_{cck} \end{cases} \quad (18)$$

Where a_1 and a_{-1} are the forward and backward reaction constants respectively. Note that this equation will be applied instead of (17) only for molecules that have been specified as receptors, which is set as a gene property.

The kinetics of binding between receptors and membrane tethered ligands are implemented differently, since the binding happens in the interface between different cells and thus receptor and ligand are located in different nodes. For that reason the receptor-ligand complex is not represented by a single regulatory molecule, but by two. If ligand *l* is expressed in node *i* and receptor *k* is expressed in node *j*, then the receptor-ligand complex will be represented by the ligand bound form *o* in node *i* and receptor bound form *p* in node *j*. The rate of change of *o* and *l* in nodes *i* and *j* respectively is:

$$\begin{cases} S_{io} = S_{jp} = a_1 g_{il} g_{jk} - a_{-1} g_{io} g_{jp} \\ a_1 = r_{klo} = r_{lkp} \\ a_{-1} = r_{kol} = r_{lpk} \end{cases} \quad (19)$$

1.4.5. Molecule degradation:

In addition all molecules have a basal degradation rate. Its rate per time is:

$$M_{ik} = \mu_k g_{ik} \quad (20)$$

Where μ_k is the intrinsic rate of degradation of molecule *k* (a model genetic parameter).
Enzyme dependent degradation (in the sense of a molecule promoting the degradation of another molecule, i.e: protease) can be implemented by defining a regulatory molecule *k* corresponding to a molecule labeled for degradation and having a second regulatory molecule (the enzyme catalyzing degradation) capable of transforming other molecules into this molecule *k*. This molecule might then have a large μ_k.

1.4.6 Diffusion (Fig. 2d).

Diffusion is implemented as transfers of molecules between nodes (including ECM nodes). This transport follows Fick's second law of diffusion:

$$\frac{\partial q}{\partial t} = -D\nabla^2 q \quad (21)$$

Where *q* is concentration of a molecule, *D* is the diffusion coefficient of that molecule and $\nabla^2 q$ is the second derivative

of the concentration in 3D space. We calculate transfers of matter between pairs of nodes. Since we only make calculations in the nodes diffusion is essentially discrete (although non-uniformly) and this equation is roughly approximated by :

$$O_{ik} = D_k \sum_{j=1}^{n_v} \left(\frac{g_{ik} - g_{jk}}{d_{ij}} \right) \quad (22)$$

Where g_{ik} is the amount of molecule k in node i, t is time, D_k is the diffusivity coefficient of molecule k, n_v is the number of nodes within the maximum radius of diffusion from node i and d_{ij} is the distance between node i and j. Both this distance and n_v depend on how nodes are arranged in space. The maximum radius of diffusion is two times the maximal p^{ADD} in the embryo in a given iteration multiplied by M_{DIF}, a model parameter. This ensures an optimal accuracy even if there are changes in the sizes of the nodes in the embryo over time.

We only consider diffusion between nodes that are closer than the maximum radius of diffusion. The amount of molecules interchanged between two nodes in each instant of time decreases with the distance between these two nodes. Thus, after some distance this diffusion becomes negligible. Increasing this distance exponentially increases the number of node pairs to be considered, thus making the calculations very expensive computationally for a small gain in accuracy. M_{DIF} is a model parameter that will usually take values between one and few node radii.

Since diffusion is only calculated between existing nodes, it cannot happen within empty cavities (without cells but filled with fluid) in the embryo, such as blastocoels. To ensure effective diffusion of molecules across those cavities, ECM nodes need to be added in order to fill them. In growing embryos those cavities will most likely grow, or decrease, so ECM should be actively secreted and/or degraded by cells (see section 5).

The same node neighborhood used for mechanical interactions is used to determine which nodes will transfer molecules between them (see Section 2.1)

In some cases we might want to simulate a developing system with open boundary conditions, where diffusive molecules would be lost through the borders. In those cases, the nodes making the boundary of the system can be set as boundary nodes, p^{BOR} node property, and then they will tend lose intercellular diffusing molecules at a rate:

$$O_{ikboundary} = D_k \left(\sum_{j=1}^{n_v} \left(\frac{g_{ik} - g_{jk}}{d_{ij}} \right) - g_{ik} \right) \quad (23)$$

1.4.7 Total amount of a molecule in a node:

In summary, the rate of change of the amount of gene product k in node i is*:*

$$\frac{\partial g_{ik}}{\partial t} = Q_{ik} - M_{ik} + O_{ik} \quad (24)$$

For other molecules this is:

$$\frac{\partial g_{ik}}{\partial t} = S_{ik} - M_{ik} + O_{ik} \quad (25)$$

1.5. Regulation of node properties

1.5.1 Node properties:

Most node properties have already been described when describing mechanical forces. See section 4.7 for a list and summary of those. Each of these values can be modified by the amounts of specific regulatory molecules in a node. Each element e_{lm} in the E-matrix describes the effect of regulatory molecule m on node property l. The value of node property l at time t in node i is then:

$$p_i^l(t) = \Phi\left(p_i^l(0) + (1 - p_i^{DIF}) \sum_{k=1}^{n_g} e_{lk} \; g_{ik} \right) \quad (26)$$

Where $p_i^l(t)$ is the value of node property l in node i at time t and $p_i^l(0)$ is the value of that node property l in node i when the node was created (this is in the initial condition or when the node first arose through growth). The Φ, as in equation (16), function ensures that node properties can become very small (or zero) but not negative. p_i^{DIF} is the degree of differentiation in node i (differentiation slows down changes in nodes). The amount of change in node properties is then related to how much of the molecules regulating these properties there is in a node and how strongly they regulate them, as specified in each element e in the E matrix. For simplicity this regulation is supposed to be instantaneous compared with the rate at which nodes move or with the rate at which regulatory molecules are catalyzed. Equation (25) applies to all node properties except for p_i^{EQD} and p_i^{DIF} that are explained later.

1.5.2 Node and cell differentiation:

In this model cell differentiation is defined as the process leading a cell to stop any developmental cell behavior or cell movement. Thus the moment when all cells in the embryo are differentiated marks the end of the developmental process in our model. The level of differentiation in a node depends on the expression of certain regulatory molecules, its rate of change is:

$$\frac{\partial p_i^{DIF}}{\partial t} = \psi\left(\sum_{m=1}^{n_g} e_{ma} g_{im} \right) \quad (27)$$

Where a is the index of the column in matrix E specifying how regulatory molecules regulate differentiation and $\psi(x)$ is a function that is 1 if x>1 and it's equal to x if x<1. As stated in equation (26), the effect of regulatory molecules on node properties is diminished by the level of differentiation of the node, being 0 when the differentiation level reaches 1. At the moment all nodes reach a differentiation level of 1, the developmental process has ended and thus the simulation stops.

1.5.3 The regulation of node radii:

The p^{EQD} of a node is the sum of four other node properties that correspond to four different cell processes.

$$p_i^{EQD} = p_i^{COD} + p_i^{GRD} + p_i^{PLD} + p_i^{VOD} \quad (28)$$

The value of p^{EQD} in each node is updated in each iteration according to the values of these other four node properties. The first term is coming from node active contraction (due to myosin and related molecules), the second is coming from cell growth and apoptosis, the third from cell mechanical plasticity and the fourth from cylinder volume conservation. Having p^{EQD} determined by four independent terms allows contraction, growth, plasticity and volume conservation to occur at the same time in a cell. For example it is important that parts of the cell can contract while the cell is growing (and that would not be possible if growth and contraction would act directly on p^{EQD} since one growth would increase p^{EQD} and contraction would decrease it: so there may not be much change overall). Cell contraction is realized when a regulatory molecule regulates negatively the node property p^{COD}. Since contraction is happening in the nodes, cells may have contraction in only part of its nodes, as it is necessary in a number of developmental processes such as in invagination by apical cell contraction. p^{COD} is calculated as in equation (26) above. The other terms are explained in the following sections and when explaining cell growth.

In addition, the model includes a minimum and a maximum for any p^{EQD}, these are model parameters M_{EMI} and M_{EMA}, but those can be set arbitrarily small or large (so that they have no effect on the model dynamics).

Special rules also apply to p^{ADD}. Any decrease or increase in p^{EQD} during a time step is also applied to p^{ADD} so that the difference between p^{EQD} and p^{ADD} does not change because of changes in p^{EQD}. This way the two node properties defining their effective size change together.

1.5.4 Node and cell plasticity:

Cells are not totally elastic. They are also viscous, as explained in section above, and can also accommodate incoming pressures by actively or passively changing the cytoskeleton. This is specially relevant in the case of the highly packed cells encountered in epithelia. With plasticity (that can be totally deactivated by setting logical model parameter L_9 to zero) epitehlial nodes accommodate to compression from other nodes by reducing their p^{EQD}. For each node *i* compression is calculated as the mean difference between the distance to its neighbors and the equilibrium distance to them:

$$Z_i = \frac{\sum_{j=1}^{n_v} \left(d_{ij} - d_{ij}^{EQD} \right)}{n_v} \quad (29)$$

Where n_v is the number of neighbors in contact with *i* and d_{ij}^{EQD} is as in equation (2). If Z_i is smaller than 0 then the node is under compression. Then the change in p_i^{PLD} per unit time is:

$$\frac{\partial p_i^{PLD}}{\partial t} = p_i^{PLA} Z_i \quad (30)$$

Where p_i^{PLA} is a node property (that can be directly regulated by regulatory molecules) specifying how plastic node *i* is.

1.5.5 Volume conservation in cylinders:

The two nodes that compose an epithelial cylinder are mostly independent from one another, nonetheless they together represent a single subcellular element and thus certain mechanical deformations in one node may affect the other node. For instance, if an epithelial cell contracts its apical surface (for example by means of actomyosin activity on the cell cortex) the volume of the apical half of the cell would decrease and, by volume conservation, the basal half should increase in volume. This kind of deformation can also occur if an epithelial sheet bends passively (from forces generated somewhere else than in the bending part; Lane et al. 1993). In this case, cells can accommodate by adopting slightly wedged shapes.

Epithelial cylinder volume conservation is implemented in the model as transfers of volume between the apical and basal nodes of the cylinder, depending on the level of compression and/or tension there is on each side. In order to make effective the transfer of volume between sides p^{VOD}, is defined as a component of p^{EQD} due to volume conservation of the cylinder, that will tend to correct any deviations from the equilibrium volume due to deformation. Its rate of change over time is:

$$\frac{\partial p_i^{VOD}}{\partial t} = p_{v\,i}^{VOC} \left(\frac{p_i^{GRD} + p_j^{GRD} - p_i^{EQD} - p_j^{EQD}}{2} \right) \quad (31)$$

Where *j* is the other node belonging to the same cylinder as *i*. This is essentially the average difference between p^{EQD} and p^{GRD} in a cylinder multiplied by the node property p_i^{VOC} ($p_i^{VOC}=p_j^{VOC}$). Note that the sum of p^{GRD} (the contribution coming from growth) from the two nodes as an equilibrium volume of the cylinder.

1.5.6 Spatial fixation of nodes.

The model allows to simulate an embryo part or organ and consider the rest of the embryo in a simpler way. This is done by simulation only the organ and having some special conditions in its boundary to represent the rest of the embryo. For example, if the system is expanding it should feel a restorative force in the boundaries, due to the resistance to compression of the tissues beyond the system.

This is accomplished by setting the node property p^{FIX} to either 1 (elastic fixation) or 2 (complete fixation) in the nodes in the borders. A node with p^{FIX} of 1 will be tied to an elastic spring anchored to the position of that same node at time 0 of the simulation and with an equilibrium length of 0. Thus, an additional force component will act on these

nodes,

$$\vec{f}_{Fi} = p_i^{KFI}\left(\vec{r}_i(t) - \vec{r}_i(0)\right) \quad (32)$$

Where p^{KFI} is a node property that determines the elastic constant of the spring in the border that multiples the vector going from the actual position of node *i* to the position it had at time 0. A node with a p^{FIX} of 2 will be completely fixed on space, thus making the boundaries of the system totally rigid.

1.5.7 Summary of node properties:

Common to all types of node:

1- Intercellular adhesion: p^{ADH}
2- Intracellular elasticity: p^{YOU}
3- Cell compressibility to nodes from the same cell: p^{REP}
4- Cell compressibility to nodes from a different cell: p^{REC}
5- Filopodia extensibility: p^{DMO}
6- Filopodia unstability: p^{MOV}
7- Node plasticity: p^{PLA}
8- Degree of differentiation: p^{DIF}
9- Equilibrium radius: p^{EQD}
10-Contration component of p^{EQD} : p^{COD}
11-Growth component of p^{EQD} : p^{GRD}
12-Plasticity component of p^{EQD} : p^{PLD}
13- Adhesion radius: p^{ADD}
14- Amount of stored ECM: p^{ECM}
15- Fixation of node in space: p^{FIX}
16- Elastic constant of fixation: p^{KFI}
17-Open boundary node: p^{BOR}

Only for epithelial nodes

18- Rotation force component resistance: p^{ERP}
19- Radial force component resistance: p^{EST}
20- Apico-basal elasticty: p^{HOO}
21- Apico-basal equilibrium distance p^{EQS}
22-Volume conservation component of p^{EQD} : p^{VOD}

1.6. Cell behaviors

1.6.1 Cell shape change and contraction.

Cell morphology is determined in this model by the size and relative position of the nodes composing a cell. Thus, cell morphology can change due to passive processes, such as deformation by mechanical stresses, or due to active processes, like genetically regulated contraction or expansion of nodes within the cell.

Cell contraction occurs as explained in section 4.3. Contraction by changes in p^{EQD} is a way for cells to produce intrinsic forces than then can spread and affect neighboring cells. Note that changes in p^{EQD} can be both increases and decreases and that even decreases will induce forces if the shrinking nodes are bound to other nodes (as it would often be the case).

1.6.2 Cell polarization and internal cell asymmetries (Supplementary Fig. 5):

Epithelial cells are by definition polarized in the apical-basal axis but can also have a polarization in the plane of the epithelium (what is often call planar cell polarity (Simons & Mlodzik 2008). Mesenchymal cells can also be polarized. In our model the polarization of a cell h is described by a 3D vector, $\hat{P}_h^{POL}$, a cell property. This vector arises from the asymmetrical distribution of regulatory molecules within the nodes in a cell. First, a polarization score s_{hi} is calculated for each node i in cell h:

$$s_{hi} = \sum_{m=1}^{n_g} c_{ma} g_{im} \quad (32)$$

Where a is the index of the column in the C matrix corresponding to the effect of regulatory molecules on cell polarity (see Supplementary Fig. 5). The polarization vector is then

$$\vec{P}_h^{POL} = \sum_{i=1}^{n_h} (s_{hi} - s_{hc}) \vec{r}_i \quad (33)$$

Where s_{hc} is the score of the node closest to the centroid of cell h, n_h is the number of nodes in cell h and $\vec{r}_i$ is the position vector of node i. This is simply an average of each node position weighted by its score, compared to that of the most central node. This vector is then divided by its module to find the polarization vector itself (the unit vector $\hat{P}_h^{POL}$). Polarization, thus, arises from asymmetries in the distribution of some molecules within the cell. These asymmetries can be present already in the initial conditions or arise during development. In development, and in our model, these asymmetries can arise by signaling between cells (or between cells and the environment) that do not lead to changes in gene expression. For example, a cell can secrete a growth factor that then, by diffusion, reaches some nodes in a neighboring cell with high concentration and some, more distant nodes of the same cell, with a lower concentration. If this signal elicits a signal transduction that promotes the catalysis of some regulatory molecule (for example the phosporylation of some protein) then there would be a gradient in the amount of this regulatory molecule within the cell (as in Supplementary Fig. 6). Signaling without changes in transcription are implemented in the model by extracellular signals activating the catalysis of specific regulatory molecules (without transcription) that then reproduce this gradient inside the cell. Note this is more difficult to produce if the signal leads to transcription since transcription occurs in the nucleus and from there diffuses in all directions within the cell (the gradient is thus from the nucleus irrespective of the direction of the gradient of the extracellular signal).

1.6.3 Cell growth (Supplementary Fig. 6):

Cell growth is implemented as a progressive addition of nodes within a cell. This can happen only if the nodes within a cell are not too compressed, this is if:

$$\sum_{i=1}^{n_h} Z_{ih} < M_{MCO} \quad (34)$$

Where n_h is the number of nodes in cell h. Z_{ih} is the level of compression on node i of cell h. Calculated fro equation (28). New nodes are added one at a time and with a small size M_{MID}, a model parameter. Then, as long as the

nodes have a p^{EQD} smaller than M_{MAE}, a model parameter, they will grow at a rate:

$$\frac{\partial p_i^{GRD}}{\partial t} = \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mb} g_{im} \quad (35)$$

Where n_h is the number of nodes belonging to cell h and c_{mb} is the effect of regulatory molecule m on cell growth (column b in the C matrix). Only when all nodes in the cell have p^{GRD} equal to M_{MAE} a new node is added to cell h. Even though growth is a process at a cell level, there is only one node increasing at a time, thus the rate of growth of that node depends on the amount of growth-inducing molecules located throughout all the cell (more than one node per iteration can be added if the logical model parameter L_{16} is set to 1). The node properties of the new node are set as the initial node properties of a random node from the same cell. Gene and molecule expression in the new node is set the same value than the node closest to it (this way the the smoothness of spacial molecular gradients is not perturbed by cell growth). Depending on where the new nodes are put in space we distinguish between non-polar and polar cell growth.

1.6.3.1 Non-polar cell growth:

In mesenchymal cells the position of the new node is chosen at random within the boundaries of the cell. In the case of epithelial cells, a new cylinder is added in a random position within the boundaries of the cell, but its orientation in the apical-basal axis is the same as that of the cylinders closest to it (to the three closest ones or to all the cylinders in a cell if there is less than three cylinders in the cell).

1.6.3.2 Polar cell growth:

Polar growth occurs with a probability:

$$P_{polar\,h} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{md} g_{im} \quad (36)$$

Where d is the index of the column in the C matrix corresponding to the effect of regulatory molecule m on the probability of polar growth (see Supplementary Fig. 5).

The position of the new node along the polarization axis is determined by finding the node which is farthest from the cell centroid and at the same time closest to the direction of polarization of the cell. This is done by calculating the dot products between the polarization vector, $\vec{P}_h^{POL}$, and the vector connecting the centroid with each node. The node giving the largest dot product will determine the direction in which the new node will be added. The new node is added at a position in the line between the cell centroid and this node and at a distance that is 80% of the distance between the centroid and this node.

1.6.4 Cell division (Supplementary Fig. 7).

1.6.4.1 Symmetric cell division:

Cell division is implemented by splitting an existing cell into two new daughter cells. In symmetric division both daughter cells inherit roughly the same number of nodes.

The triggering of division in a cell depends on two factors: the progression of the cell cycle and the number of nodes in the cell. Progression of the cell cycle is specified by the cell property P^{PHA} and can take values from 0 to 1 (being 1 when division takes place). The rate of increase in P^{PHA} on any cell h is:

$$\frac{\partial P_h^{PHA}}{\partial t} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{md} g_{im} \quad (37)$$

Where d is the index of the column in the C matrix corresponding to the effect of regulatory molecules on cell phase progression. n_h is diving the sum to ensure that just having more nodes does not affect phase progression. This is thus the sum of the contributions of all the regulatory molecules affecting that cell behavior in all the nodes of a cell. In addition, for a cell to divide it is also required that it has at least P^{MIN} nodes (a cell property). Also if a cell has more than P^{MAX} (another cell property) it divides irrespectively of its phase (P^{MAX} can, however, be set to be arbitrarily large). As we

later explain the values of these cell properties, as all other ones, can be modified by regulatory molecules.

In nature, it is often the case that the plane of division is normal to the longest axis of the cell, what is commonly referred as Herwig's rule (Minc et al. 2011), or normal to the polarization axis of the cell (as specified in section 5.2). The longest cell axis (Hertwig vector: vector $\vec{P}_h^{HER}$, Supplementary Fig. 8) is calculated by means of a 3D linear regression of nodes' positions. The actual division vector (the vector normal to the plane of division) is calculated as a weighted average of the Hertwig and polarization vectors:

$$\vec{P}_h^{\mathrm{DIV}} = (1-w_h)\hat{P}_h^{POL} + w_h \hat{P}_h^{HER} \quad (38)$$

The weighing factor for any cell *h*, w_h, is calculated as a function of the concentration of regulatory molecules in cell *h* that affect this weight. This is:

$$w_h = \frac{1}{1+\sum_{i=1}^{n_h}\sum_{m=1}^{n_g} c_{me}\, g_{im}} \quad (39)$$

Where *e* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on polarized cell division. Thus, if W_h =1 the division vector is equal to the Hertwig vector and if W_h=0 it is equal to the polarization vector.

The actual plane of division is normal to the division vector $\vec{P}_h^{\mathrm{DIV}}$ and is passes through the centroid of the cell. This plane splits the cell in two and nodes in one side are assigned to one daughter cell and nodes in the other to the other. The former nucleus of the cell loses its identity (it becomes a normal node), and after the division the most central node (the node closest to the new centroids) of each new cell is chosen as the nucleus.

1.6.4.2 Asymmetric cell division.

In asymmetric division the size of the two daughter cells is different (one daughter cell has more nodes than the other). In this case, the position of the division plane along the division vector, $\vec{P}_h^{\mathrm{DIV}}$, does not pass by the physical center of the cell (the centroid), but depends on the spatial distribution of certain regulatory molecules within the cell. Each node *i* in cell *h* gets a score determined by summing all the molecules affecting the asymmetry of division:

$$s_{hi}^{ASY} = \sum_{m=1}^{n_g} c_{mf}\, g_{im} \quad (40)$$

Where f is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on asymmetric cell division. Then the division plane is placed at the point in the axis defined by $\vec{P}_h^{\mathrm{DIV}}$ where the sums of the scores of nodes at each side of the plane are equal. Thus, the more skewed the distribution of those molecules, the more asymmetric is the cell division. If the gene product distribution is uniform then the plane of division appears on the centroid of the cell.

If the division is very asymmetric and the cell has not a very regular shape then daughter cells with isolated nodes can be produced (that is nodes in a cell not having physical contact with each other). Since this situation is biologically unrealistic outcome of cell division, the physical integrity of potential daughter cells is checked before cell division. If a daughter cell has unconnected nodes, the division plane is moved again to a position closer to the centroid, until the two new cells have all their nodes connected.

1.6.5 Cell death (Supplementary Fig. 8).

Cell death or apoptosis is implemented in this model as inverted cell growth, that is when a cell is dying nodes start to decrease in size (p^{GRD}) until a minimum size is reached, then the node is deleted from the simulation. When all nodes belonging to a cell disappear then the cell also does so. The rate of decrease of p^{GRD} and p^{ADD} are equal to:

$$\frac{\partial p_i^{GRD}}{\partial t} = \frac{\partial p_i^{ADD}}{\partial t} = -\sum_{m=1}^{n_g} c_{mg} g_{im} \quad (41)$$

Where *g* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on apoptosis.

1.6.6 Cell adhesion:

Cell adhesion is integrated in the mechanical part the model (see Section 2). Each node has a basal adhesivity plus the one given by the expression of adhesion molecules, which depends on the affinity of the adhesion molecules expressed in each node. As discussed in Section 2, this includes also the possibility to implement repulsion between cells.

1.6.7 Epithelial-mesenchymal transition (EMT) (Supplementary Fig. 9).

EMT is implemented as a discrete transformation of an epithelial cell to a mesenchymal one. Each epithelial cylinder is converted into two mesenchymal nodes by changing the identity of the two epithelial nodes and removing the elastic spring between them. Before that, in order to keep the spatial continuity of the future mesenchymal cell, the apical and basal sides of the epithelial cell are brought closer to each other by reducing the length of the elastic springs. The transition is regulated by a cell property P_h^{EMT} that is progressively increased due to gene expression,

$$\frac{\partial P_h^{EMT}}{\partial t} = \frac{1}{n_h} \sum_{m=1}^{n_g} c_{mk} g_{im} \quad (42)$$

Where *k* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on EMT. n_h is diving the sum to ensure that just having more nodes does not affect P_h^{EMT}. When P_h^{EMT} reaches a value of 1, the transition is realized.

1.6.8 Secretion:

1.6.8.1 ECM secretion (Supplementary Fig. 10):

ECM is represented as free spheric nodes, which can be secreted by any type of cell provided that there is expression of molecules regulating its secretion. Given that ECM nodes in the model represent finite amounts of large fibrous extracellular molecules like proteoglycans or collagen, those have to first accumulate within the cell before being secreted as one ECM node. The rate at which those products accumulate within a node is,

$$\frac{\partial p_i^{ECM}}{\partial t} = \sum_{m=1}^{n_g} c_{ml} g_{im} \quad (43)$$

Where l is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on ECM secretion and p_i^{ECM} is the amount of accumulated ECM products within node *i*. Once p_i^{ECM} reaches a value of the model parameter M_{ECM} a node is secreted near node i and p_i^{ECM} is set back to 0. The p^{GRD} of the ECM node will be equal to M_{ECM}, correlating in a way the amount of ECM components and the volume of the node. The other components of p^{EQD} will be set to zero. The p^{REC} and p^{ADD} of a new ECM node *i* are determined at the moment of secretion depending on gene expression in the node that secreted it:

$$p_i^{REC} = \sum_{m=1}^{n_g} c_{mr} g_{i'm}$$
$$p_i^{ADD} = p_i^{EQD} + \sum_{m=1}^{n_g} c_{ms} g_{i'm} \quad (44)$$

Where *i'* is the node that secreted node *i*, *r* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecule *m* on the amount of ECM matter secreted, *s* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecule *m* on the amount of adhesive ECM matter. Note that ECM nodes can

not have p^{YOU} or p^{REP} since they do not belong to a cell.

Some regulatory molecules expressed in the cell may be secreted along the ECM, such as adhesion molecules or other compounds that are tightly bound to the ECM fibrous components. Those molecules will be transferred to the ECM node at the time it is secreted. The molecules secreted this way can not diffuse between nodes but can react with molecules that are diffusing between ECM nodes (e.g.: bind to extracellular signals). The proportion of regulatory molecule m that would be secreted with the node i and the proportion that would remain in the original node *i'* are:

$$\begin{aligned} g_{im} &= c_{mt}\ g_{i'm} \\ g_{i'm} &= (1-c_{mt})\ g_{i'm} \end{aligned} \quad (45)$$

Where t is the index of the column in the *C* matrix corresponding to the propensity of regulatory molecules to be secreted.

In epithelial cells, the new ECM node is placed at a small distance of the node that secreted it, in the direction of the apical-basal axis (either apical and basal epithelial nodes can secrete ECM). In mesenchymal cells the ECM node is placed in at a random position in the line going from the cell centroid to the node that is secreting the node (See Supplementary Fig. 10). If the cell has only one node then the ECM node is secreted in a random direction (at a short distance).

1.6.8.2 Diffusible extra-cellular signals:

The process of secretion of extracellular diffusible molecules happens when an intracellular gene form is transformed via catalytic activity into an extracellular form. The latter will immediately start to diffuse to other cells or to the ECM. Extracellular diffusible molecules can act as extra-cellular signals by binding to cell surface receptors or by affecting the mechanical properties of the ECM nodes and even mediate its degradation, acting as extracellular proteases (Shapiro 1998).

ECM degradation can be specified as a gene property in the *C* matrix. The degradation of ECM nodes is implemented similarly as in the case of cell death. The presence of a gene product specified as an extracellular proteases in an ECM node will promote a decrease in the node's size at a certain rate. When the ECM node reaches a minimum size, it disappears. The rate at which an ECM node shrinks due to protease mediated degradation is:

$$\frac{\partial p_i^{GRD}}{\partial t} = \frac{\partial p_i^{ADD}}{\partial t} = -\sum_{m=1}^{n_g} c_{mg} g_{im} \quad (46)$$

Where *g* is the index of the column in the *C* matrix corresponding to the effect of regulatory molecules on ECM degradation.

1.7. Regulation of cell properties:

In the following section we explain how the cell properties $\vec{P^{POL}}$ (cell polarity), P^{PHA} (cell cycle progression), P^{EMT} (EMT transition) are changed.

The cell centroid is also a cell property, it is simply the average of the positions of each of its nodes in 3D space (it is thus a 3D vector). This centroid is calculated in each iteration. P^{MIN} and P^{MAX} are regulated by regulatory molecules:

$$\frac{\partial P_h^{MIN}}{\partial t} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mq} g_{im} \quad (47)$$

Where q is the index of the column in the C matrix corresponding to the effect of regulatory molecules on the minimal number of nodes for cell division P^{MIN}. n_h is the number of nodes in cell h, and it is diving the sum to ensure that just having more nodes does not affect P^{MIN}.

$$\frac{\partial P_h^{MAX}}{\partial t} = \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{m=1}^{n_g} c_{mw} g_{im} \quad (48)$$

Where w is the index of the column in the C matrix corresponding to the effect of regulatory molecules on the maximal number of nodes allowed before cell division P^{MAX}. n_h is the number of nodes in cell h, and it is diving the sum to ensure that just having more nodes does not affect P^{MAX}.

1.8. Numerical integration

Differential equations are numerically integrated by the explicit Euler method or, optionally, by the explicit fourth-order Runge-Kutta method. These methods can be used with a fixed time step, with an adaptive time step or by a dynamic time step. The logic model parameters L_7,L_{10} and L_{15} specify how the numerical integrations are performed.

The adaptive step-size integration is done by the standard step-doubling procedure. When the integration is not adaptive the value of δ, the integration time step, can be either set constant (when logical model parameter L_7 is set to 1) or dynamic (default, L_7=0) over time, depending on the maximum node movement length at each time step. The dynamic δ value is calculated as:

$$\delta_t = \frac{M_{DDA}}{|\vec{r}_{max}|} \quad (49)$$

Where δ_t is the value of δ at time t, $|\vec{r}_{max}|$ is the length of the longest movement vector in the system at that time and M_{DDA} is a model parameter that specifies the value of δ_t when $|\vec{r}_{max}|$ is unity. This ensures that when changes in node positions occur very fast, large $|\vec{r}_{max}|$, the calculations are done with higher accuracy (smaller δ_t). In that sense M_{DDA} also specifies how accurate the calculations are (higher accuracy when low M_{DDA} values).

We further control the value of δ by setting a maximum value of δ, model parameter M_{DMA}. If δ is larger than that value, δ is set equal to that value. At the same time if δ is smaller than the model parameter M_{DMI} (a model parameter) then δ is set equal to that value. Real time increases by δ_t per time iteration.

The numerical integrations can also be done with a constant δ by setting the logical parameter L_{10} to 1. In that case δ is equal to M_{DMI}.

1.9. Model parameters:

1.9.1 Numerical model parameters:

These have all been explained already but here we summarize them:

-M_{TEM}: Temperature analog, this is how likely are noisy movements that are energetically unfavorable (see section 2.5).

-M_{NOI}: Proportion of nodes to which noise is applied in each iteration. If delta is dynamic this proportion is weighted by δ_t (see 2.5)

-M_{MCO}: Maximal compression allowed in a cell to allow growth in it (see 5.3).

-M_{EMI}: Minimal p^{EQD} allowed.

-M_{EMA}: Maximal p^{EQD} allowed.

-M_{MAE}: p^{EQD} all nodes in a cell should have before adding a new node.

Implementation model parameters. These are parameters controlling the numerical implementation of the model, this is the accuracy of the model. They have no biological meaning as such.

-M_{DIF}: Maximum radius of diffusion. By default its value is equal to 2. Values larger than that have a negligible effect on accuracy and largely decrease the speed of the model.

-M_{MID}: p^{EQD} given to the new added nodes by growth. Any value that is small compared to the average p^{ADD} of nodes (or compared to M_{MAE}) would produce the same model dynamics.

-M_{ECM}: Amount of extra-cellular matrix that has to accumulate in a node before an ECM node is secreted. This essentially controls how much ECM there needs to be for the model to consider that ECM as a node.

-M_{DMI}: Minimum δ. In any case, the numerical integration step may not be below this value. The lower this value is, the more accurate are model calculations but the slower the model would be. This parameter is only meaningful if delta is dynamic (L_7=0, that is the default).

-M_{DDA}: Accuracy of the numerical integration. This parameter is only meaningful if δ is dynamic (L_7=0, that is the default).

-M_{DMA}: Maximal value of δ allowed. This parameter is only meaningful if δ is dynamic (L_7=0, that is the default).

-M_{MNN}: Maximal number of nodes any node can interact with. If there is more than that number the program crashes. These is no optimal way to avoid that effect, since these neighbors need to be stored in a temporary matrix and there are system restrictions in the size of those. In addition, there is no way to predict how many nodes a node will interact with since this is a result of model dynamics. If this value is large the program would run slower.

-M_{AMX}: In the mechanical interaction between two cylinders, the minimum curvature in order to apply the epithelial rotational force and the epithelial surface tension force (see Section 2.4).

-M_{RMA}: Maximum node length of movement when δ is dynamic (L_7=0).

-M_{MAN}: Maximum number of nodes allowed in a simulation. Only applicable when (L_2=1).

-M_{GAB}: Sets the size of the exclusion sphere when using the Gabriel method to build the node's neighborhood (only when L_3=1).

-M_{DFE}: Maximum p^{EQD} allowed for an epithelial node due to deformation (See section 4.3)

There also some model parameters that only apply if some logical model parameters are set to 1. This is they have no meaning in the default version of the model but are relevant for some alternative versions of the model activated by some logical model parameters. They are explained in the next section.

1.9.2 Logic model parameters:

Here we describe their non-default values and how these change the functioning of the model. All logical model parameters are by default set to 0 and that defines the canonical version of the model. By setting any of those parameters equal to 1 different variations of the model can be activated (all of them are only small variations). These options can be manually altered in the input and output files or can be edited through the gene network viewer.

L_1: If set to 1 the model considers that each mesenchymal cell is made of a single node and each epithelial cell is made of a single cylinder. The initial conditions have to be designed consistent with that (each mesenchymal cell should have a single node and each epithelial cell a single cylinder).

In this case slightly different rules apply in some cell behaviors:

In the case when cells are composed of one node/cylinder, internal asymmetries cannot arise (except in epithelial cells, but only in the apical-basal axis), thus polarization has to be determined by the molecules present in the surrounding cells. In this case, the cell will polarize in the direction of a tissue-level molecular gradient. Thus, the s scores (equation 32) are not calculated for the nodes within the same cell (since there is only one, or two) but for the nodes of the neighboring cells. Then, $\vec{P}^{POL}$ is calculated as in equation in the non single node case using the s scores of the neighboring cells.

Cells are normally supposed to double their size before dividing. Thus, in the case when cells are composed of one node/cylinder they only need to add one node/cylinder before dividing. In this case, cell growth happens at the same time as cell division, by adding a new node/cylinder and then splitting the cell in two (See section 5.4).

In the case when cells are composed of one node/cylinder, they doesn't need to reach a minimum size nor a maximum size in order to divide. When P_h^{PHA} reaches the value of one, a new node/cylinder is added in the direction of the division vector and the cell is split in two, each node/cylinder belonging to a different cell.

The division vector, as explained above, depends on the weighted sum of $\vec{P}^{POL}$ and $\vec{P}^{HER}$, and in this case $\vec{P}^{HER}$ has to be calculated differently, since the shape of the cell cannot be determined by the relative position of its nodes (since there is only one). In this case we assume the shape is determined by the distances between the cell and its neighbors. If the distance between the cell and some of its neighbors is longer in a certain direction we assume that the cell is elongated in that direction. Thus, for simplicity, $\vec{P}_h^{HER}$ will be equal to the vector connecting the two neighboring cells which are farthest from one another.

Chemotaxis cannot be implemented as explained above when cells are composed of one node/cylinder. In this case, cells may move in a random fashion, but biased by the direction of cell polarization. Expression of certain regulatory molecules may determine the strength of this bias,

$$\vec{r}_i^{noise} = \vec{X} + P_h^{POL} \sum_{m=1}^{n_g} c_{mb} g_{im} \quad (50)$$

Where $\vec{r}_i^{noise}$ is the movement vector of node i by noise, $\vec{X}$ is a random unit vector with a spheric distribution, P_h^{POL} is the cell's polarization vector, c_{mb} is the element of the *C* matrix specifying the effect of regulatory molecule *j* on chemotaxis and g_{im} is the expression of regulatory molecule *m* on node *i*.

L_2: If set to 1 stops the simulation after a threshold number of nodes are reached. This number is an optional model parameter, M_{MAN}.

L_3: If set to 1 the Gabriel algorithm to determine which nodes interact with each other is used. This allows nodes to screen the interaction between other nodes (as explained in above). This option is often used combined with L_{11} set to 1.

L_4: If set to 1 it disables epithelial surface tension.

L_5: If set to 1 stops the simulation if any node gets a p^{EQD} or p^{ADD} value three times larger than its original value.

L_6: If set to 1 forces apoptosis of all cellular nodes that are not interacting with any other one after M_{TAL}, a model parameter, iterations.

L_7*:* If set to 0 it uses explicit Euler method for the numerical integration, if set to 1 it uses explicit fourth-order Runge-

Kutta for node movement and explicit Euler for the rest of equations. If set to 2 it uses explicit Runge-Kutta for node movement and regulatory molecule equations.

L_8: If set to 1 stop the simulation once all cells are fully differentiated.

L_9: If set to 1 it allows node plasticity.

L_{10}: If set to 1 fixed value δ is used instead of dynamic δ.

L_{11} : If set to 1 calculates the neighborhood between nodes by using the delaunay tesselation (see SEC above). Otherwise neighborhood is established simply by distance.

L_{12}: If set to 0, the mechanical interaction between two cylinders from the apical/basal side is calculated as described in Section 2. If set to 1 it is calculated as if they were two mesenchymal nodes (see Section 2)

L_{13}: If set to 1 volume conservation in cylinders is implemented.

L_{14}: If set to 1 the diffusion of p^{EQD} components is allowed. This is there is diffusion of p^{GRD} , p^{COD} and p^{PLD} between mesenchymal nodes from the same cell and between epithelial nodes from the same side. This simply reflects that since the cytosol and membrane of different parts of a cell communicate there would be a natural re-distribution of matter between cell parts, this is between nodes. This re-distribution process is analogous to a diffusion in the sense that the flux of matter would be from nodes with higher values on those properties to nodes with lower values. The diffusivity of that process is model parameter, M_{DID}.

L_{15}*:* If set to 1 uses adaptive step-size and fourth-order explicit Runge-Kutta numerical integration for node movement. If set to 2 it uses it also for regulatory molecules equations.

L_{16}: If set to 1 the simulations are run by allowing more than one node to be added per cell per iteration due to growth.

L_{17}: If set to 1 and dynamic δ is used returns the control to the user in the embryo display after the number of real time units that have been run (by default returns the control after the specified number of iterations and not of real time units). This option is only valid if the model is run with the user graphical interface (see GNOMO_Sim user manual).

L_{18}: If set to 1 noise is implemented without considering energies. This option is highly unrecommended unless the user knows what he/she is doing or unless debugging is being performed.

L_{19}: If set to 1 runs the model only by energy biased noise (Monte Carlo Method). This in general does not change model outcomes but makes the simulations much more slower.

L_{20}: If set to 1, epithelial nodes from one side may consider as neighbors epithelial nodes from the opposing side. By default those neighbor connections are not considered. This may be relevant when two different epithelial surfaces are close to each other, but it is not necessary when the same epithelium folds over itself, since the nodes in contact belong to the same surface.

L_{21}: By default (L_{21}=0) single element cells divide by adding a new, full sized cell next to it. When L_{21}=1 the new cell has a smaller size and then gradually grows until it has full size.

1.10. Implementation of the model in EmbryoMaker.

1.10.1 Structure of the code. The source code is written in fortran90 and is organized in different functional fortran modules. The most relevant modules are listed below:

- general.mod.f90: declarations of the main variables that are used in common by the rest of the modules. These are global model parameters and node and cell properties. The set of all node properties are declared in a derived type fortran 90 variable. The same occurs for cell properties. The main variables used by the other modules are a matrix of nodes and cells properties (one element per cell and node). Essentially the rest of the code is mostly operations on those matrices (including re-dimensioning them).

- model.mod.f90: manages the temporal progression of the developmental simulations and calls the subroutines in the neighboring, bio-mechanical, genetic and nexus modules (once per iteration with Euler and several times with Runge-Kutta).

- neighboring.mod.f90: contains the subroutines to calculate the neighbor relations between nodes.

- biomechanic.mod.f90: contains the subroutines that calculate the mechanic interactions and displacement of nodes.

- energy.mod.f90: contains the subroutines that calculate energy potentials for nodes that are used in energy-biased random movements.

- genetic.mod.f90: declares the regulatory molecules and their parameters used in a specific instance of the model. It also contains the subroutines for transcription and non-transcriptional regulation.

- nexus.mod.f03: contains the subroutines that implement the molecular regulation of node and cell properties and the calls to the cell behaviors. It also contains subroutines for some simple cell behaviors.

- growth.mod.f90: contains the subroutines that implement cell growth.

- death.mod.f90: contains the subroutines that implement cell death or apoptosis.

- mitosis.mod.f90: contains the subroutines that implement cell division.

- ecm.mod.f90: contains the subroutines that implement secretion of extracellular matrix.

- single_node.mod.f90: contains certain subroutines that are used in the case cells are composed of one node.

- pinta.mod.f90: This file contains two modules: A view_modifier module to control how the embryo is seen (rotation, zooming, sectioning, etc...) and a function_plotter module that contains all the subroutines that draw nodes and controls the menu. This latter module is the one including the OpenGl and glut calls.

- editor.mod.f90: contains the code required to manually edit the embryo.

- ic.mod.f90: contains a set of subroutine for simple initial conditions.

- initial.mod.f90: contains several initialization subroutines.

- io.mod.f90: contains the hard-disc input/output subroutines.

-OpenGl_gl.f90, OpenGl_glu.f90 and OpenGl_glut.f90 define fortran interfaces for the OpenGl, GLU and glut functions, and have been taken from the f03gl project (http://www-stone.ch.cam.ac.uk/pub/f03gl/index.xhtml)

1.10.2 Input/Output format. EmbryoMaker and NetworkMaker use a custom I/O format. The same file written by the program as output can be read as input file as well. It is basically a text file listing all the model parameters and variables that are used by the software, including node positions and gene expression levels. The names of the parameters and variables are indicated in the file, so it is possible to edit the file manually. In that sense, both the editor tools of EmbryoMaker and NetworkMaker can be used to edit Input/Output files with a more intuitive graphic interface. By default, as explained in the manual, EmbryoMaker writes all the output files from a given run into a folder with a number (a different number for each run) within a folder called output.

1.11 References.

Bagnard, D., Lohrum, M., Uziel, D., Puschel, A. W., & Bolz, J. (1998). Semaphorins act as attractive and repulsive guidance signals during the development of cortical projections. *Development*, *125*(24), 5043-5053.

Brizzi, M. F., Tarone, G., & Defilippi, P. (2012). Extracellular matrix, integrins, and growth factors as tailors of the stem cell niche. *Current opinion in cell biology*, *24*(5), 645-651.

J Delile, R Doursat, N Peyriéras. Computational Modeling and Simulation of Animal Early Embryogenesis with the MecaGen Platform , in Computational Systems Biology, 2nd Edition ed. A Kriete and R Eils, 2014

Graner, F., & Glazier, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. Physical review letters, 69(13): 2013-16.

Lane, M. C., Koehl, M. A., Wilt, F., & Keller, R. (1993). A role for regulated secretion of apical extracellular matrix during epithelial invagination in the sea urchin. Development, 117(3), 1049-1060.

Minc, N., Burgess, D., & Chang, F. (2011). Influence of cell geometry on division-plane positioning. *Cell*, *144*(3), 414-426.

Newman, T. J. (2005). Modelling multi-cellular systems using sub-cellulars elements. Math. Biosci. Eng. 2. 611-622.

Reinitz, J., & Sharp, D. H. (1995). Mechanism of< i> eve</i> stripe formation.Mechanisms of development,49(1), 133-158.

Purcell, R.M. 1977. Life at low Reynolds number.American Journal of Physics.

Salazar-Ciudad, I., García-Fernández, J., & Solé, R. V. (2000). Gene networks capable of pattern formation: from induction to reaction–diffusion. Journal of theoretical biology, 205(4), 587-603.

Salazar-Ciudad, I., Jernvall, J., & Newman, S. A. (2003). Mechanisms of pattern formation in development and evolution. Development, 130(10), 2027-2037.

Shapiro, S. D. (1998). Matrix metalloproteinase degradation of extracellular matrix: biological consequences. *Current opinion in cell biology*, *10*(5), 602-608.

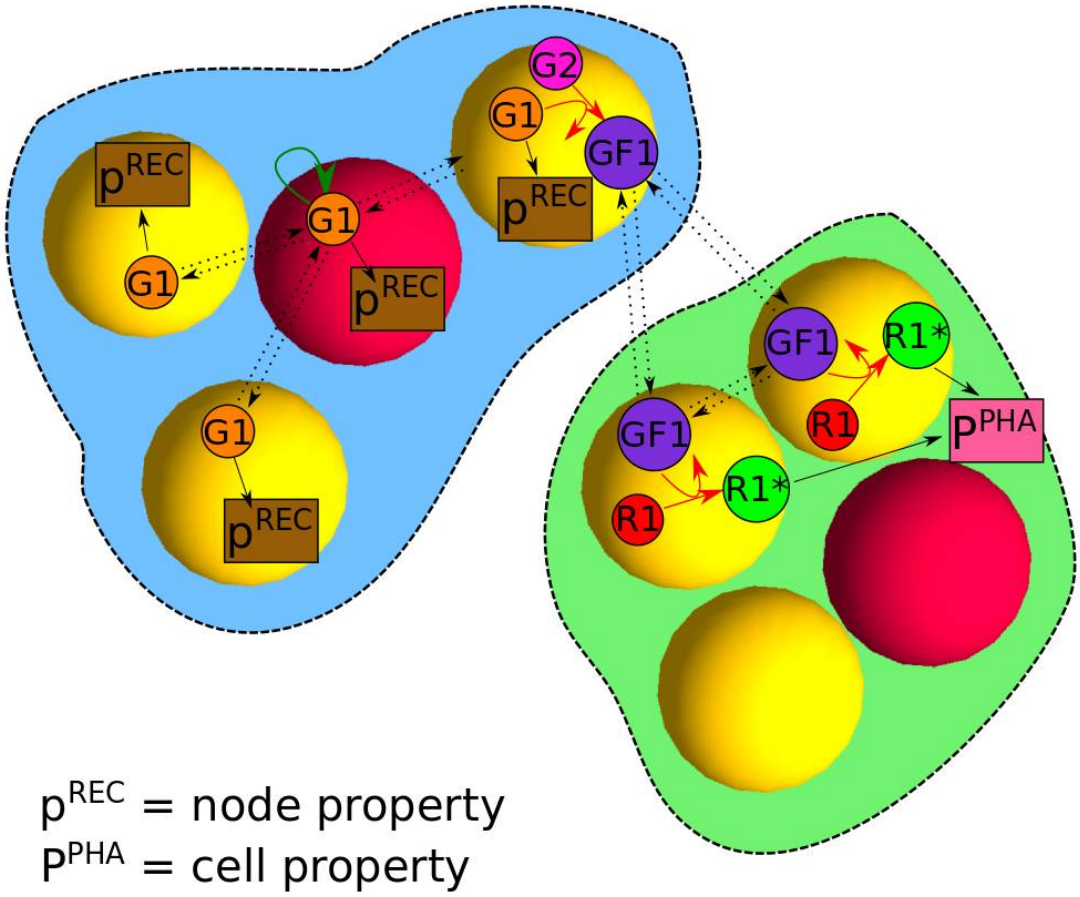
Simons, M., & Mlodzik, M. (2008). Planar cell polarity signaling: from fly development to human disease. *Annual review of genetics*, *42*, 517.

Wang, H. U., & Anderson, D. J. (1997). Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth. *Neuron*, *18*(3), 383-396.

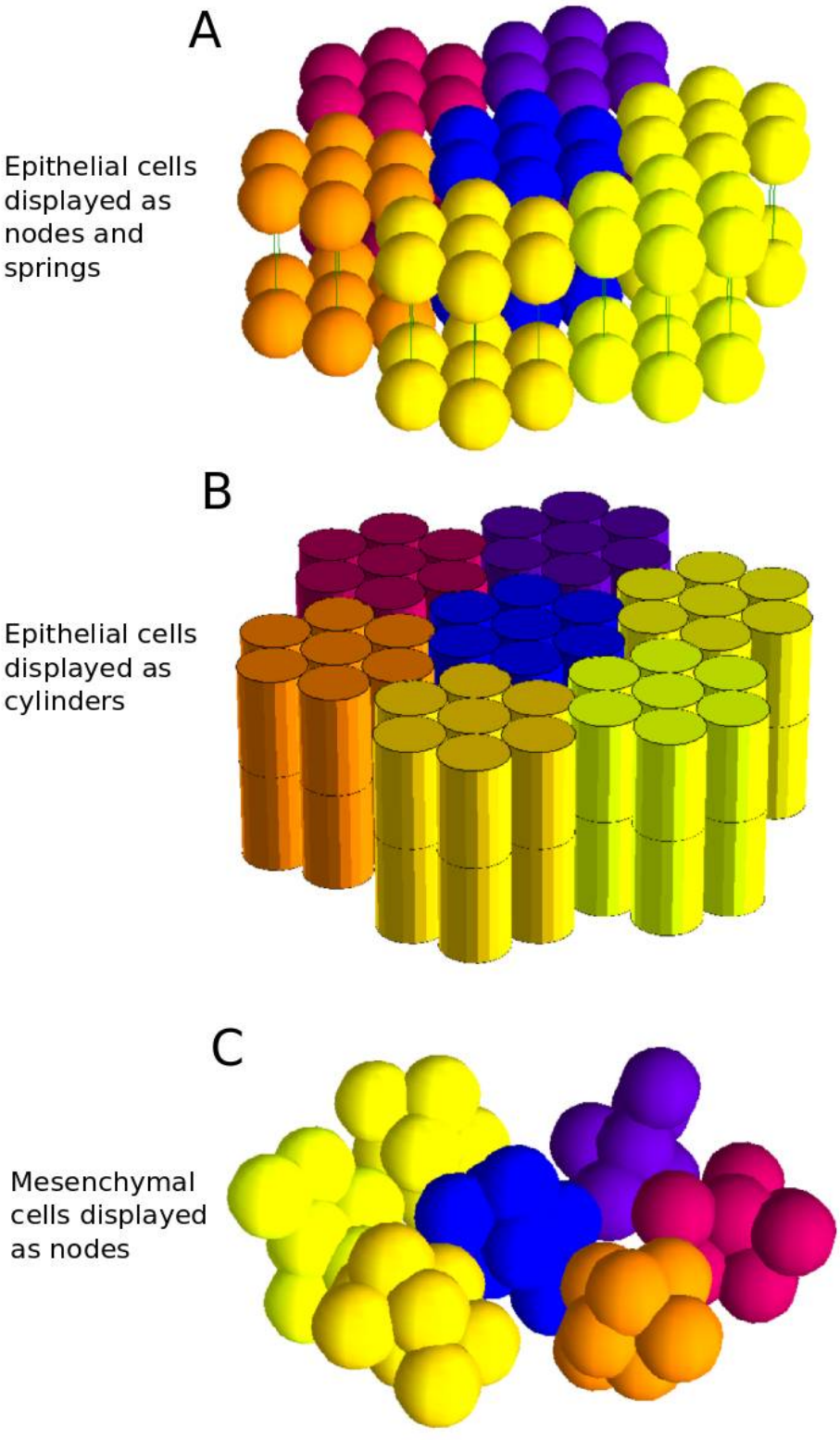
2. Detailed explanation of the developmental mechanism shown in figure 6.

Figure 7 shows an example of a combination of basic developmental mechanisms (contraction, polar growth, ECM secretion and hierarchic inductive mechanisms). The figure shows how from the network depicted in the left and the initial conditions in the center left (time zero) the patterns in the right of the figure will arise over time. It is important to note that nothing else than this network and the initial conditions are specified to the model (there are no pre-patterns or changes in the rules of the model over time). The developmental patterns shown in the right of the figure simply arise from model dynamics. These pattern transformations can be explained qualitatively. The initial pattern consists of a hollow spherical epithelium (in the simulation shown in the figure each element is a cell) in which one cell expresses transcription factor 1 (TF1) (all other regulatory molecules are either not expresses or expressed homogeneously in all cells in the initial condition). TF1 promotes an epithelial-mesenchymal transition (EMT) and thus the single cell expressing that gene detaches from the epithelium and moves randomly in the interior of the blastula. TF1 activates the transcription of TF6, a gene that promotes cell motility, cell proliferation and extracellular matrix secretion. TF1 promotes the production and secretion of growth factor 1 (GF1). As a result, while this EMT is still taking place, GF1 reaches the nearby epithelial inner surface. All epithelial cells express a receptor for GF1 (R1) at the same level in the initial conditions. Thus, GF1 bind its receptor in the epithelial cells that are close enough to the proliferating mesenchymal cells (daughters of the cell originally expressing TF1 and that thus also express GF1). The activated receptor (RGF1) activates the transcription of TF2. TF2 activates cell contraction in the outer surface of the epithelium (this is a decrease in the p^{EQD} in these nodes), thus mediating a slight invagination in the epithelium next to where the mesenchymal cells are. TF2 also mediates the production of a second growth factor (GF2). In this case, GF2 is only secreted in the outer surface of the epithelial that end up expressing it. GF2 binds to receptor R2 which is expressed in all epithelial cells from the initial conditions. The activated receptor RGF2 activates the transcription of TF3, which mediates the increase in size of the outer side of the epithelial cells where it becomes expressed. Since TF2 strongly inhibits transcription of TF3, this one will be expressed only around the territory where TF2 is expressed, but without overlap. This means that while contraction in the TF2 territory promotes a concavity in the epithelium, TF3 mediated expansion in the surrounding cells will have the opposite effect, promoting a convex curvature surrounding the TF2 mediated concavity. Also, since TF3 expression relies on GF2 signaling, this means that the farther from the TF2 territory (GF2 source) the lower the concentration of TF3, thus creating a gradient of TF3 along the whole epithelium, being highest close to the TF2 territory and lowest at the opposite side. TF3 also promotes cell cycle progression where it is expressed, meaning that cells close to the TF2 territory will divide more rapidly than the ones farther from it, but not the ones within the concavity. TF3 also inhibits the transcription of another transcription factor (TF5) that is homogeneously expressed in all the epithelium. Thus, TF5 forms a gradient opposite to the one formed by TF3, meaning that is lowest near the TF2 territory and highest in the opposite side. TF5 also promotes polarization of cells, meaning that cells will become polarized along the gradient formed by TF5. This means that the cells located on the opposite side of the TF2 territory will divide in the direction of this gradient, thus promoting an elongation of the whole embryo in that direction. Close and within the invagination this oriented proliferation leads to a deepening of the invagination towards the inside of the embryo.

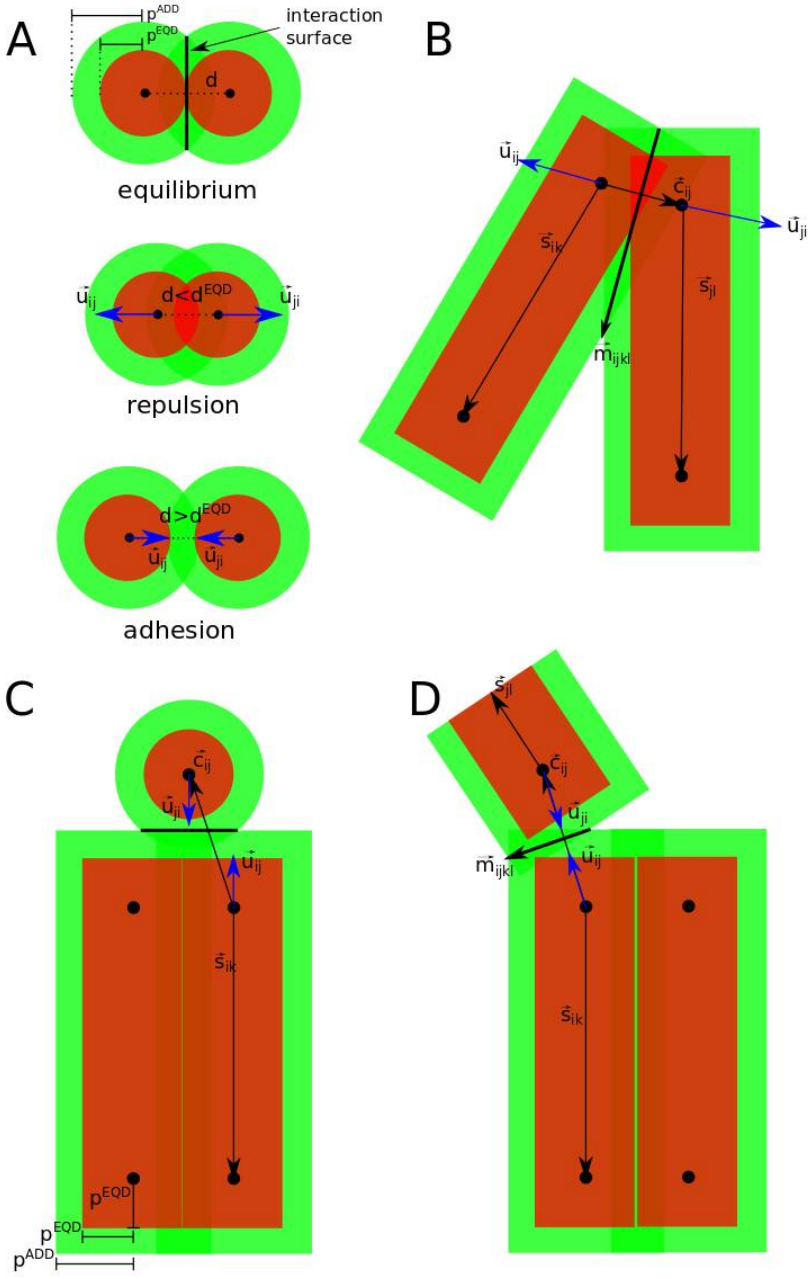
Supplementary Figure 1. Schematization of the multiscale structure of the model by and example. On the macroscale, cells are represented by sets of integrated mechanical bodies (subcellular elements) that physically interact with each other by means of different mechanic equations. The mechanic coefficients used in these equations are affected quantitatively by the amount of certain regulatory molecules present on each node. The amount of different types of regulatory molecules within each node is determined by the microscale dynamics, which include gene transcription, enzymatic reactions, spatial diffusion and molecule degradation. Here, two different cells are shown. Subcellular elements, or nodes, are represented as spheres. All nodes belonging to the same cell are enveloped by a dashed line. The red node in each cell contains the cell nucleus, thus only in those nodes transcription will take place. Regulatory molecules are indicated as colored circles with a letter and a number: G for a generic regulatory molecule, GF for a secreted growth factor, R for a receptor of secreted growth factors and R* for the activated receptor (the receptor bound to its growth factor ligand). Node properties are indicated with a lowercase p within a rectangle and cell properties are indicated with an uppercase P within a rectangle. In the figure we put the example that the P^{REC} node property and P^{PHA} cell properties are affected. Green arrows represent transcriptional regulatory interactions (see equation 4 in the main text), red arrows indicate enzymatic reactions (see equation 5 in the main text), black solid arrows represent regulation of node or cell properties (see equation 6 in the main text) and doted arrows indicate diffusion between nodes (equations 21 and 22 of the supplementary text).



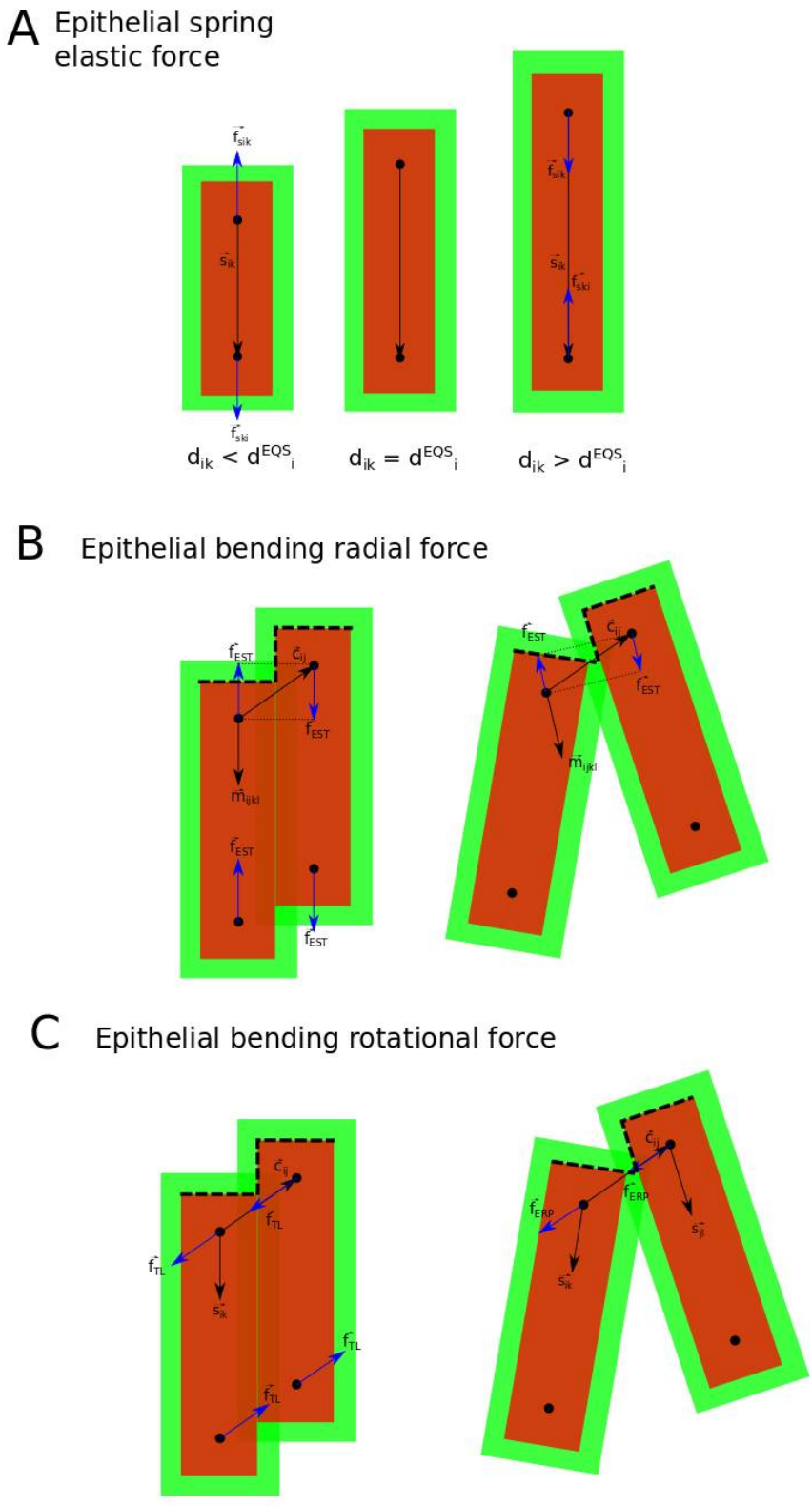
Supplementary Figure 2. Some examples of initial conditions. A, a small epithelial sheet made up of 7 epithelial cells, each one made up of 14 nodes. Spheres represent nodes and green lines represent the springs connecting two nodes forming a cylinder. Nodes of the same color belong to a single cell. B, the same as A, but displaying the form of the cylinders. C, a small group of 7 mesenchymal cells made up of 10 nodes each cell.



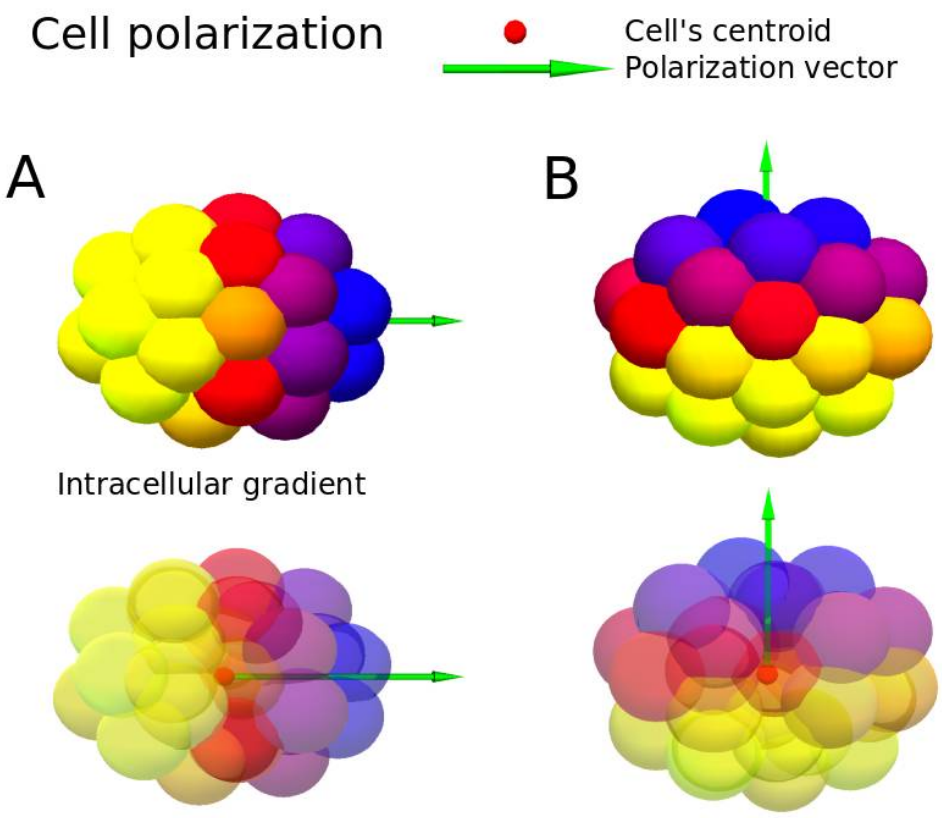
Supplementary Figure 3. Basic mechanical interactions. A, when two spheric nodes (either mesenchymal or ECM) are at a distance closer than d^{ADD} they feel either an attractive force if they are closer than d^{EQD}, or a repulsive force if they are farther than d^{EQD}. The direction of the force goes from the center of one node to the center of the other. B, D, when two cylinders interact the same repulsive and attractive forces act, but the direction of the force is always normal to the contact surface of the cylinders. C, When a spheric node interacts with a cylinder apical or basal face, the direction of the force is always parallel to the apical-basal axis of the cylinder.



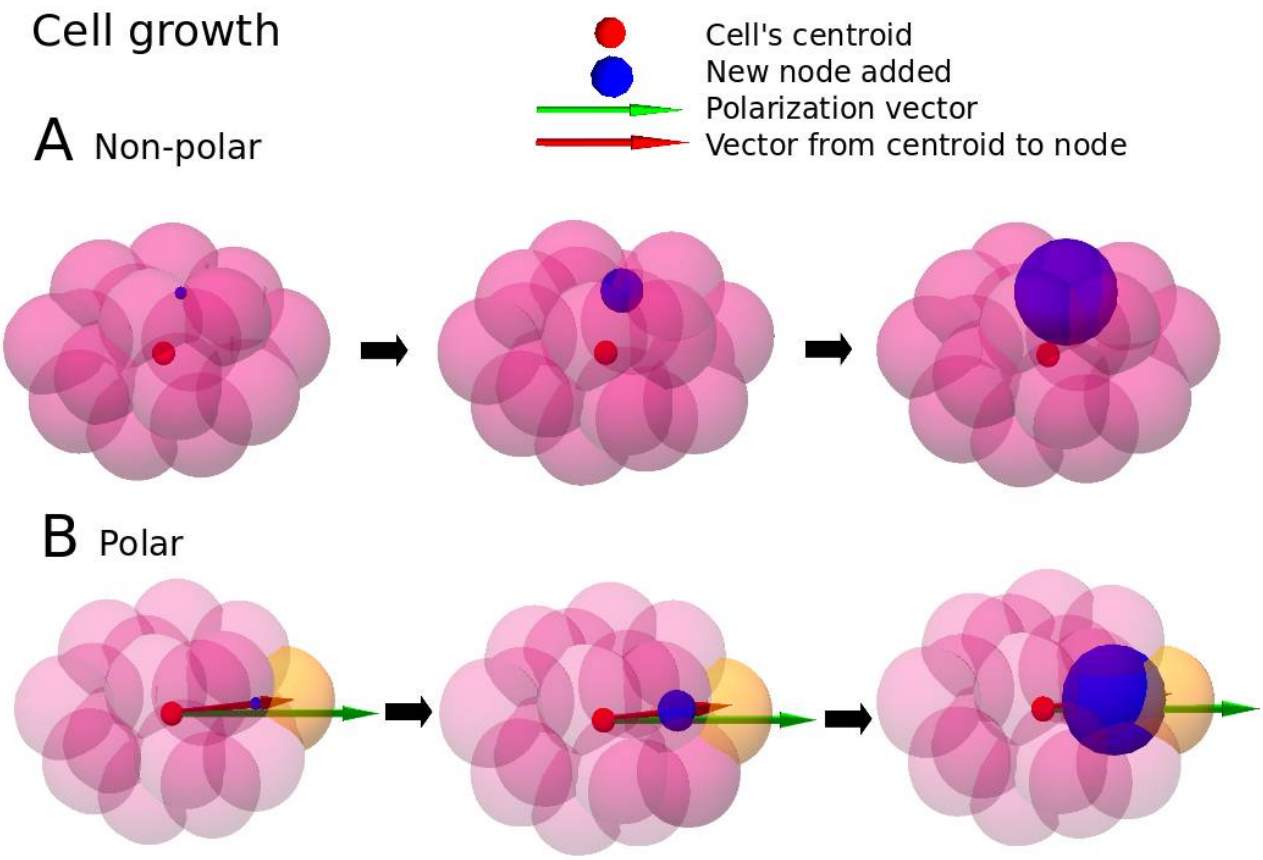
Supplementary Figure 4. Epithelial specific mechanical forces. A, the two nodes composing a cylinder are connected by an unbreakable spring. Elastic forces will always follow the direction of that spring. B, C, epithelial bending forces tend to put two cylinders in a position in which the angle between the vector connecting the two apical (or basal) nodes and the apical-basal axis is π/2. B, The bending radial force applies on a direction normal to the apical/basal surface. C, The bending rotational force applies in the direction connecting the two epithelial nodes from the same side.



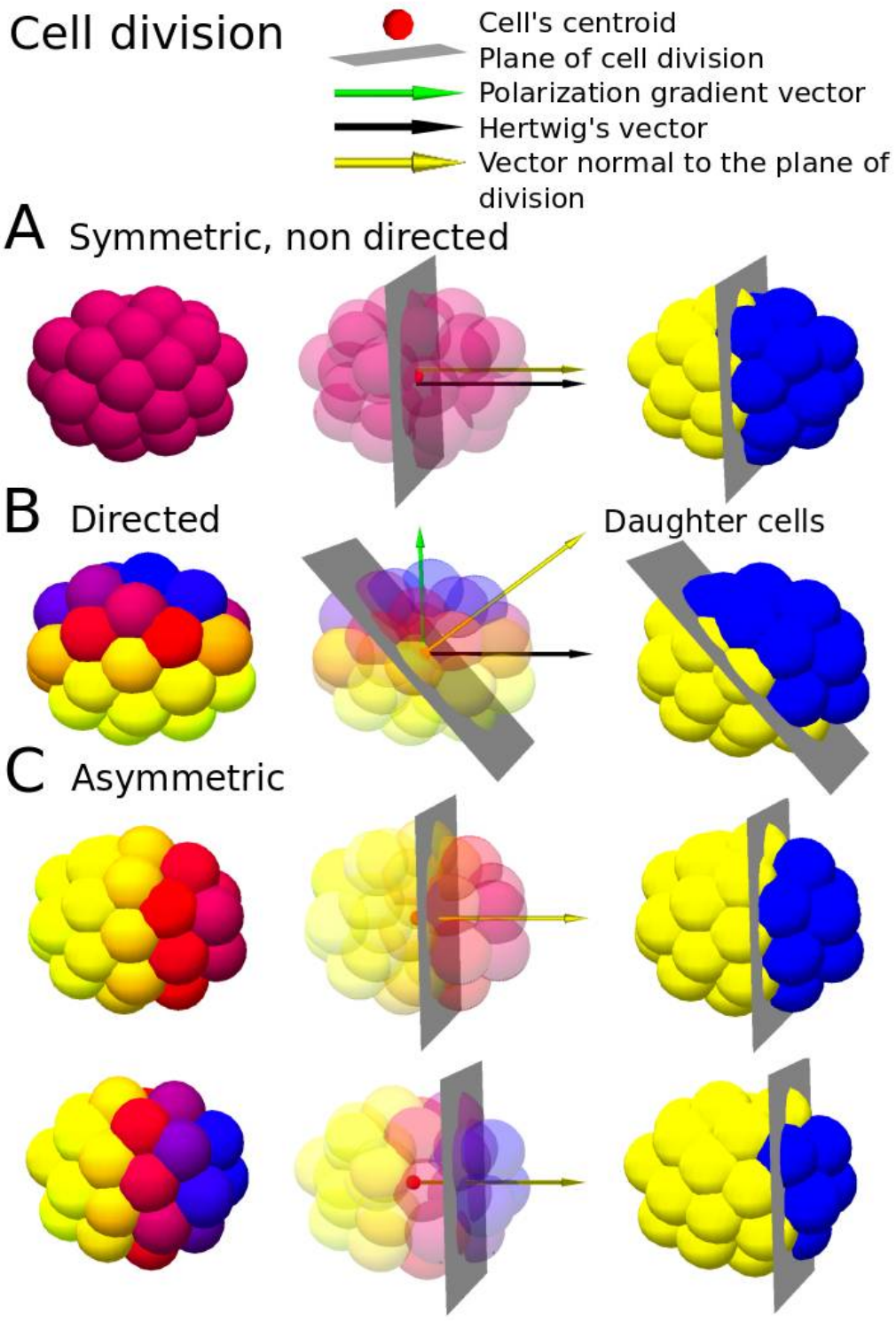
Supplementary Figure 5. Cell polarization. Polarization is calculated as the sum of node-centroid vectors weighted by the difference in molecular concentration respect the actual geometrical centroid of the cell. In other words, the polarization vector (green arrow) will mostly point towards the part of the cell where the concentration of the molecule is higher (A and B, blue shaded nodes). As B shows, polarization vector is independent of cell shape.



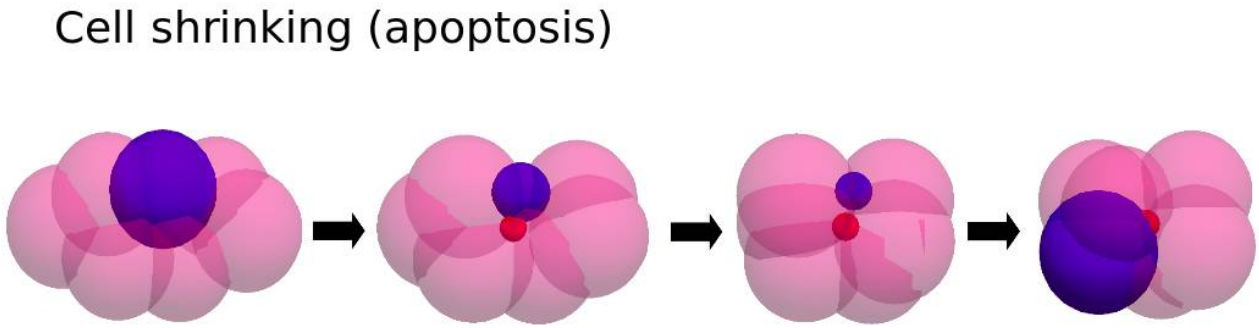
Supplementary Figure 6. Adding new nodes during cell growth. A, non polar growth. A new node (blue) is added in a random place within the cell. B, polar growth. The most external node in the direction of the polarization vector is chosen. Then a new node is added at 80% of the distance from the cell's centroid to that node. In epithelial cells the same is done, but just using the basal layer of nodes, and then completing the cylinder adding one more node to the corresponding position in the apical layer.



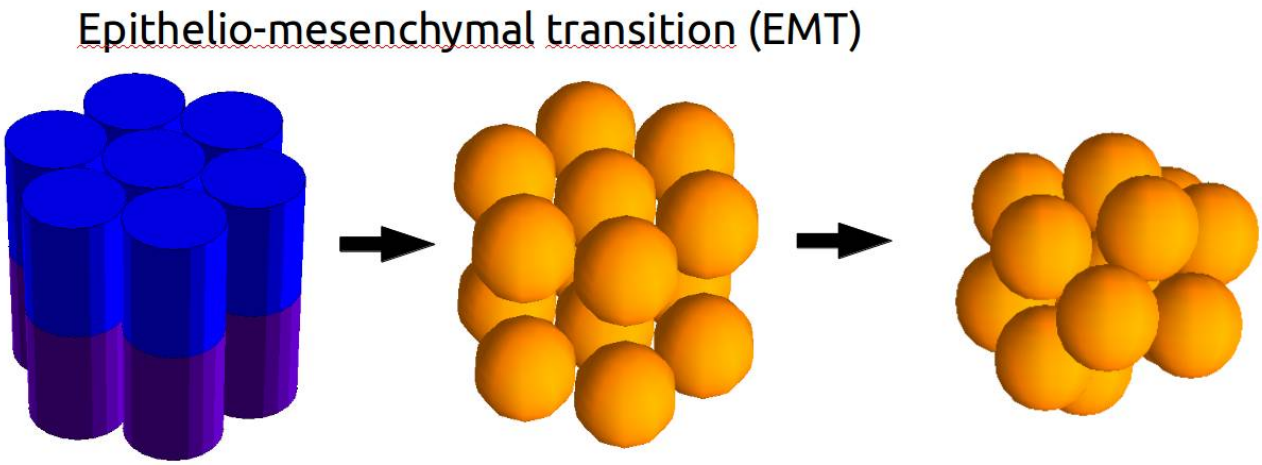
Supplementary Figure 7. Cell division. A, symmetric, non-directed. The plane of division passes through the cell's center of mass, or centroid (small red solid ball) and the vector normal to that plane (yellow arrow) is the longest axis of the cell (Hertwig's or shape vector, black arrow). B. Directed mitosis. The plane of division passes through the cell's centroid, but the direction of the plane is the weighted sum of two vectors, a cell polarity dependent vector (green arrow), and the Hertwig vector. C, in assymetric cell division, once the vector normal to the division plane (yellow arrow) has been stablished, the plane of division is set away from the centroid of the cell so one daughter cell is larger than the other. The displacement of the plane depends on the direction and intensity of a particular spatial molecular gradient within the cell. B and C can be combined to get directed assymetric mitosis. In all cases, once the plane is set, the nodes at each side of it will be assigned to a single daughter cell.



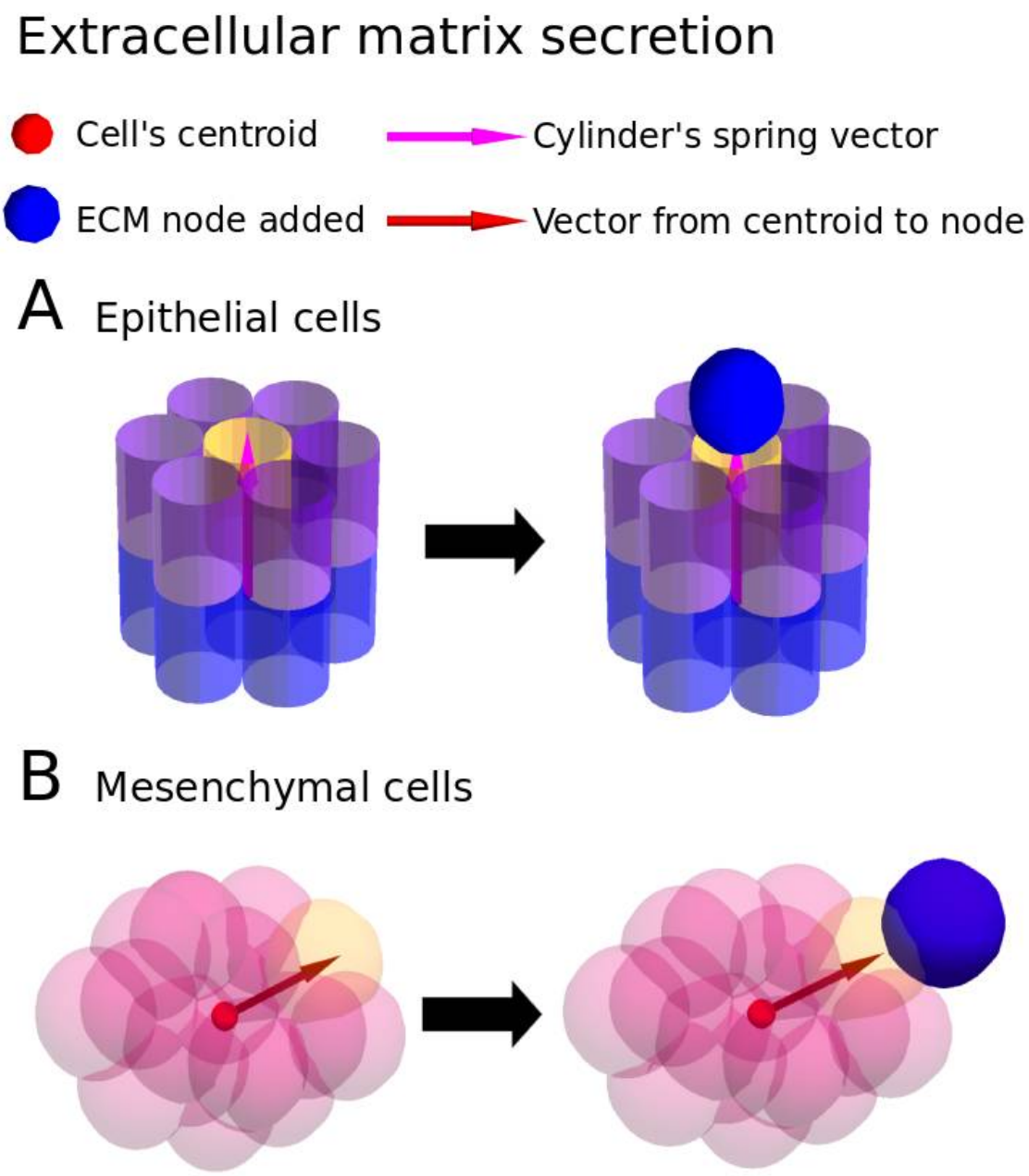
Supplementary Figure 8. Cell shrinking leading to programmed cell death (apoptosis). When there are gene products promoting apoptosis within a cell there is a continued decrease in size by shrinking one node at a time (in blue). When the node reaches a minimum size (a model parameter) it disappears and another random node begins to shrink (rightmost picture). This process keeps going until the last node in the cell disappears and so does the cell.



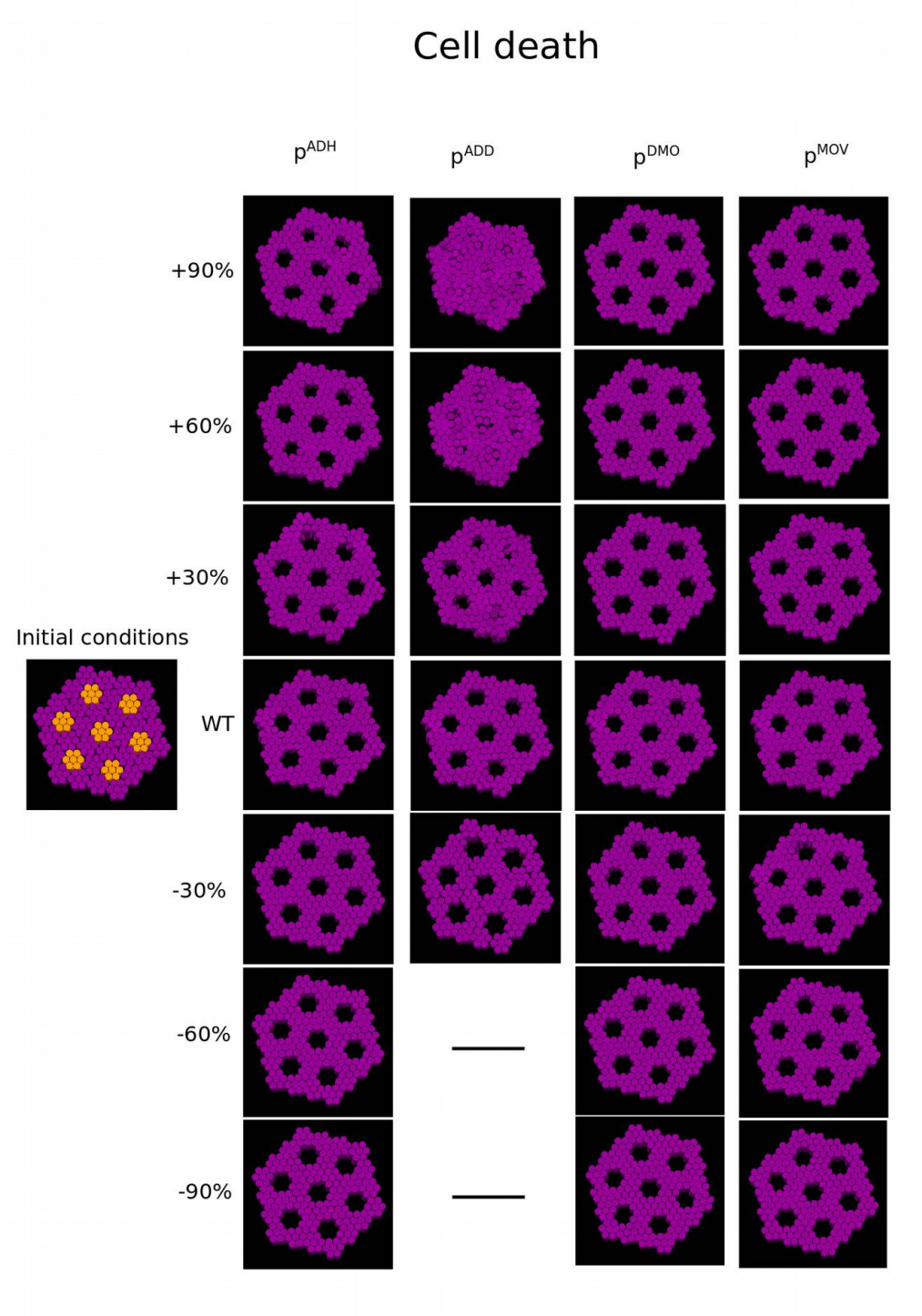
Supplementary Figure 9. Epithelial to mesenchymal transition (EMT). An epithelial cell, represented as a set of cylinders is transformed into a mesenchymal cell by turning each cylinder into a pair of spheric nodes. Since any spring or surface tension forces are exerted on the nodes once they undergo the EMT, the nodes lack the typical arrangement of epithelial cells and become more irregular (right).



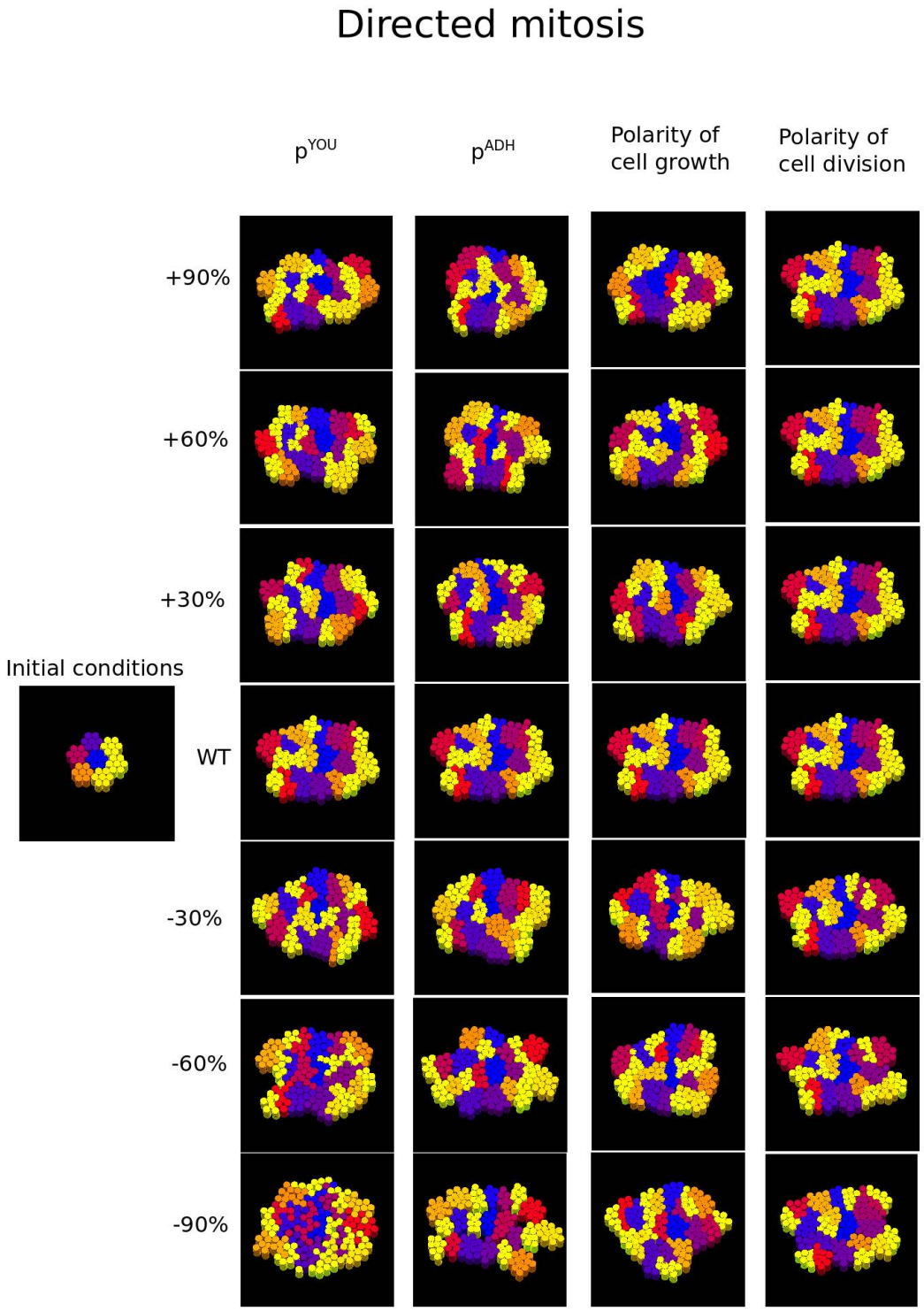
Supplementary Figure 10. Extracellular matrix (ECM) secretion. ECM is represented as free spheric nodes (depicted in blue) that can be secreted by cells. A, when an epithelial cell secretes an ECM node, it appears close to the tip of the cylinder (in yellow) where the ECM products were accumulating. B, when a mesenchymal cell secretes an ECM node, a vector from the node to the centroid of the cell is chosen (red arrow), next the node closest to that vector and farthest from the centroid (in yellow) is selected. Then, an ECM node is added close to that node and in the opposite direction from the centroid of the cell. That way, intracellular deposition of ECM is prevented.



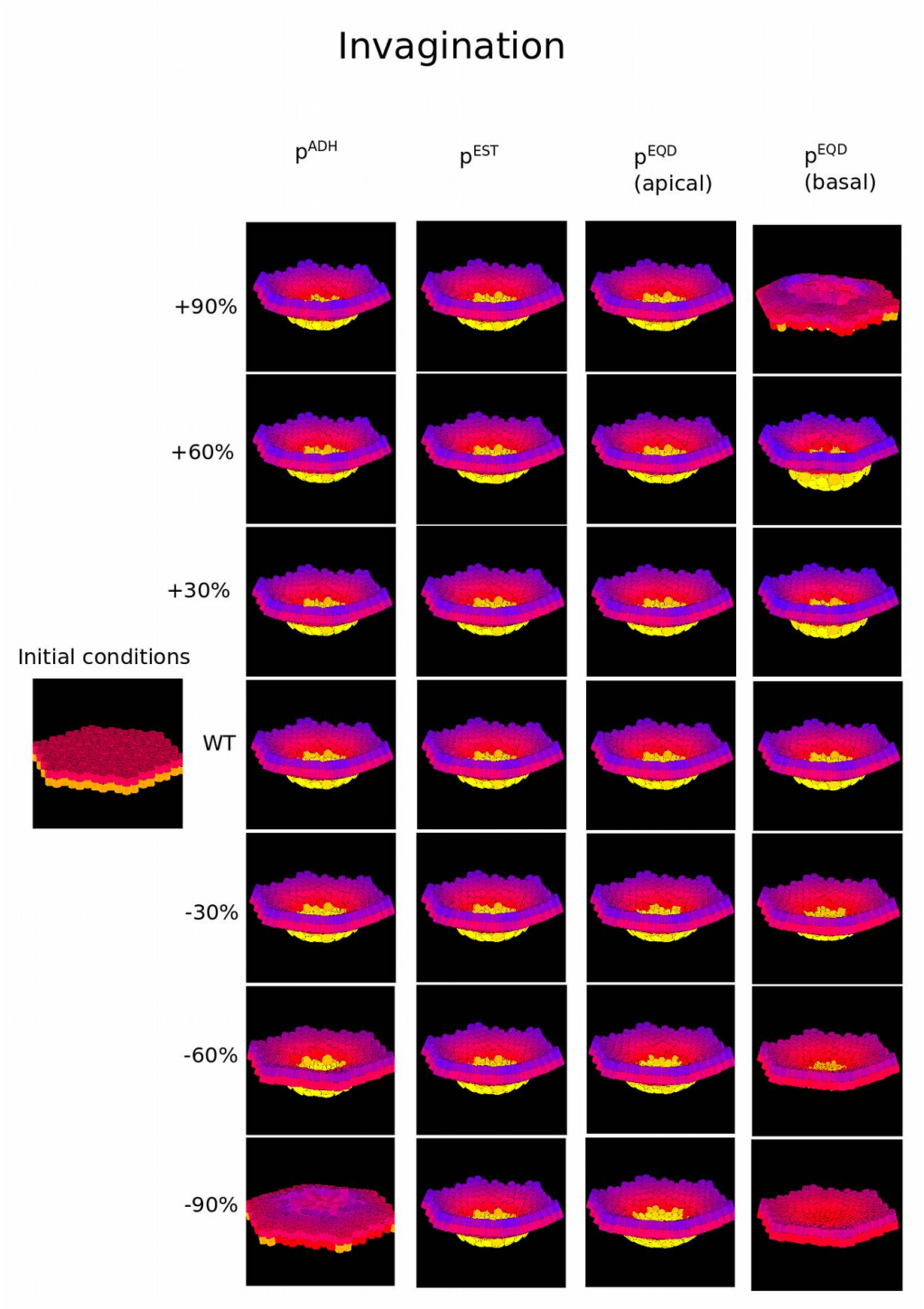
Supplementary Figure 11. Cell death genotype-phenotype map. Different simulations were run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). Seven non-adjacent cells of an epithelial sheet, each one composed of seven cylinders, express an apoptotic factor (marked in orange in the initial conditions) promoting cell death. After the cell death, the existence and shape of the resultant gaps in the final patterns depend of many parameters. For example, we show in this simplified parameter space that the remaining cells are able to close the gaps if the node's radius (p^{ADD}) is large enough as to interact with the nodes in the other side of the gap. Variations in the remaining node properties (intracellular adhesion, radius and filopodia unstability, respectively p^{ADH}, p^{ADD} and p^{MOV}) do not seem to have a strong effect on final phenotypes. In this genotype map, as well as in many following ones, the lack of final phenotypes corresponding to -60% and -90% decreases of maximum radius of interaction between nodes stems from the fact that in those cases, the radius of interaction is lower than the radius of equilibrium, thus preventing interactions between nodes.



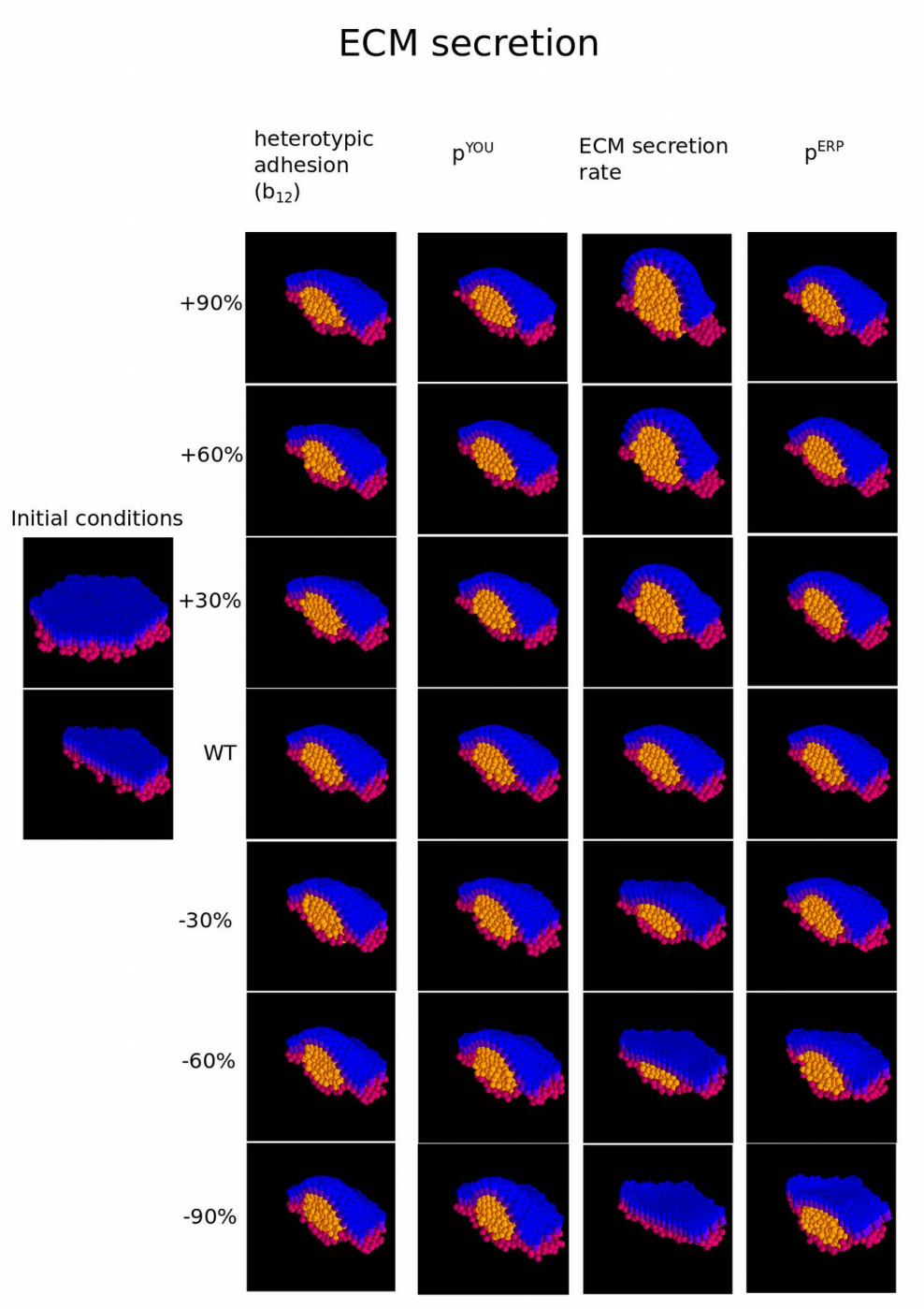
Supplementary Figure 12. Directed Growth and division genotype-phenotype map. Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). A small epithelial sheet composed of seven cells (each one depicted in a different colour) is allowed to grow in a medium displaying a chemical gradient (from left to right, not shown). When cells grow, new nodes are prefferentially added in the direction of the gradient and, when cells divide, the plane of cell division is also oriented according to that gradient. The combined effects of these mechanisms produces that epithelium gets elongated in the same direction of the gradient. This elongation is more pronounced under some parameter combinations, but fades out under others (e.g. reduced cell-cell adhesion: pADH < wild type pADH). Irregular cell shapes found in (p^{YOU}< 60%) stem from the fact that in these cases the intracellular adhesion between nodes practically equals extracellular one, so the cohesive forces keeping cell integrity vanish.



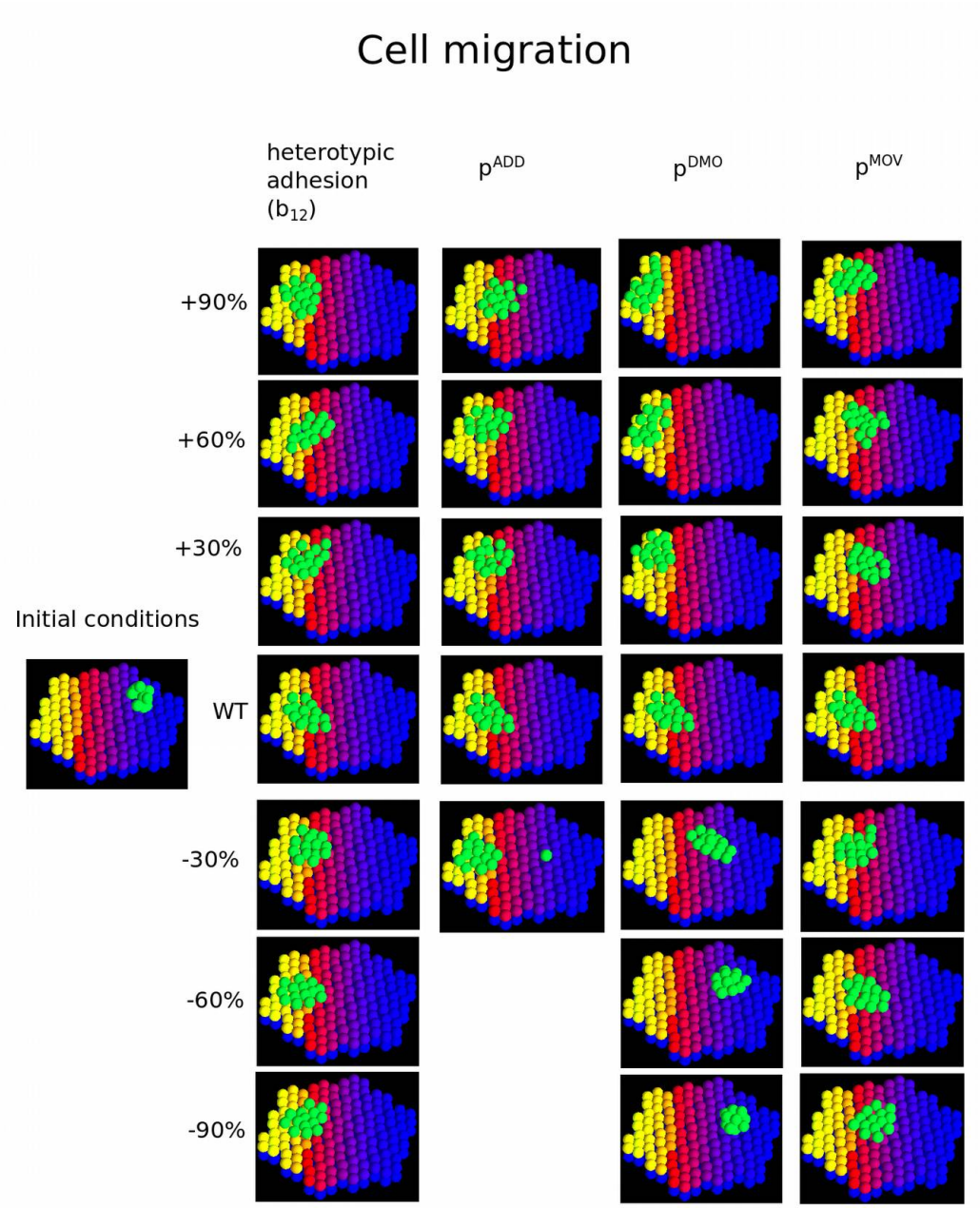
Supplementary Figure 13. Apical contraction genotype-phenotype map. Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). An hexagonal epithelial sheet invaginates when cells located in its central area perform apical cell contraction. However, the depth of such invagination varies according to many parameters (in this plot, the deeper nodes are from their initial conditions position, the more yellow they are coloured). Under some parameter combinations, invagination does not occur. This happen when apical contraction is too low ($p^{EQD} < 90\%$), or when epithelial cells lack cohesivity between them (either by decreased cell-cell adhesion ($p^{ADH} < 90\%$), or by extreme apical contraction ($p^{EQD} > 90\%$)). Radial component of epithelial surface tension (p^{EST}) and equilibrium distance (p^{EQD}) seem to have only a mild effect on apical contraction, at least for the range of values presented in here.



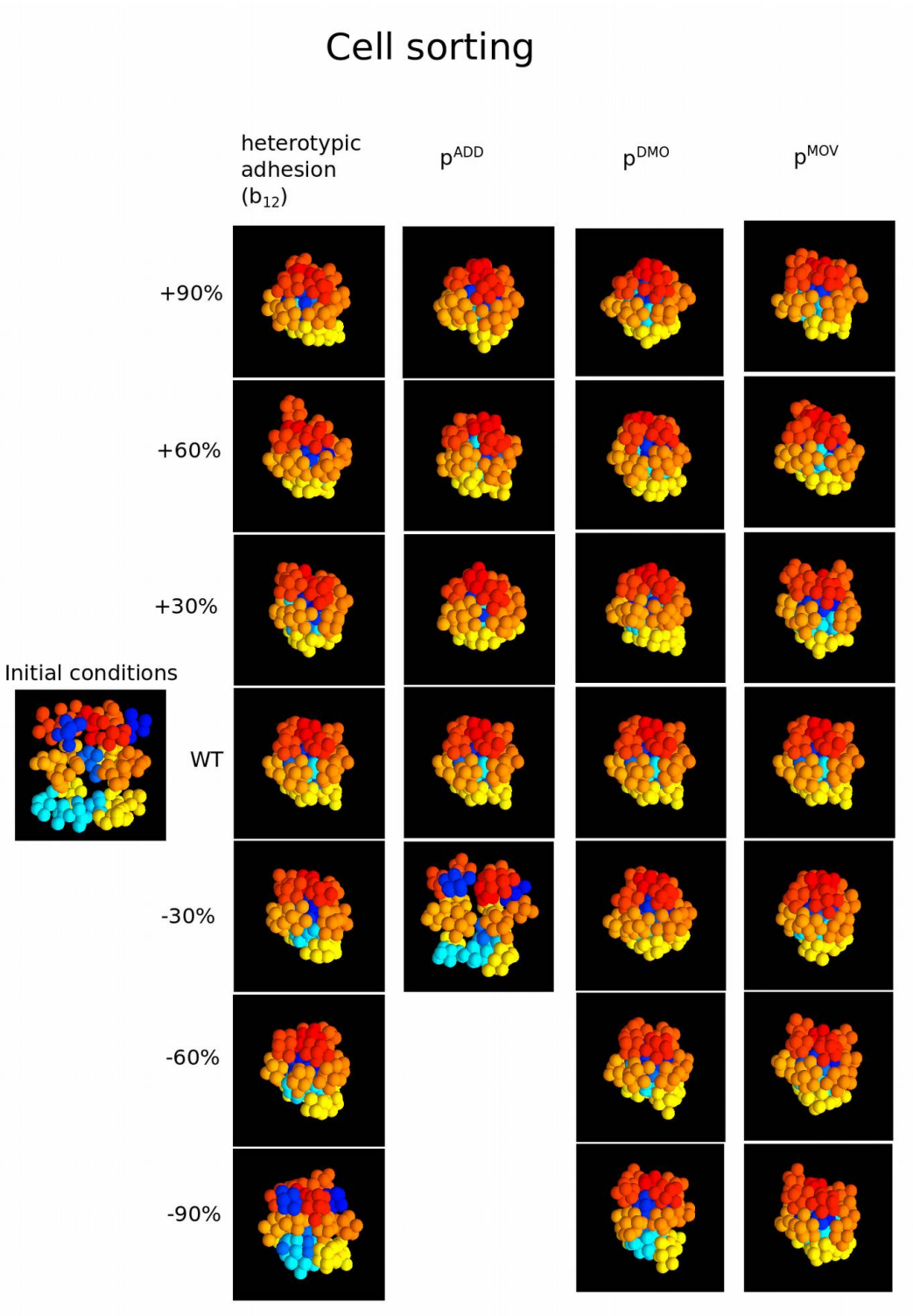
Supplementary Figure 14. Extracellular matrix secretion genotype-phenotype map. Different simulations where run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). In these simulations, extracellular matrix (orange nodes) is secreted by epithelial cells (in blue) in an extracellular space previously occupied by mesenchymal cells (pink). A section of the system is depicted for a better visualization. The mechanical forces resulting as a result of ECM secretion deform both the epithelial sheet and the underlying mesenchyme, but the relative magnitudes of such deformations depend on many parameters. When heterotypic adhesion is high, ECM nodes tend to spread between epithelial and mesenchymal cells, but they tend to form a rounded aggregate when this adhesion diminishes. Not surprisingly, deformation also strongly depends on the rate of ECM secretion, even deformation is not accomplished if the rate is too low. When the bending accommodation force (p^{ERP}) diminishes, epithelial cells become less adaptive to bending, so epithelia is only slighly deformed.



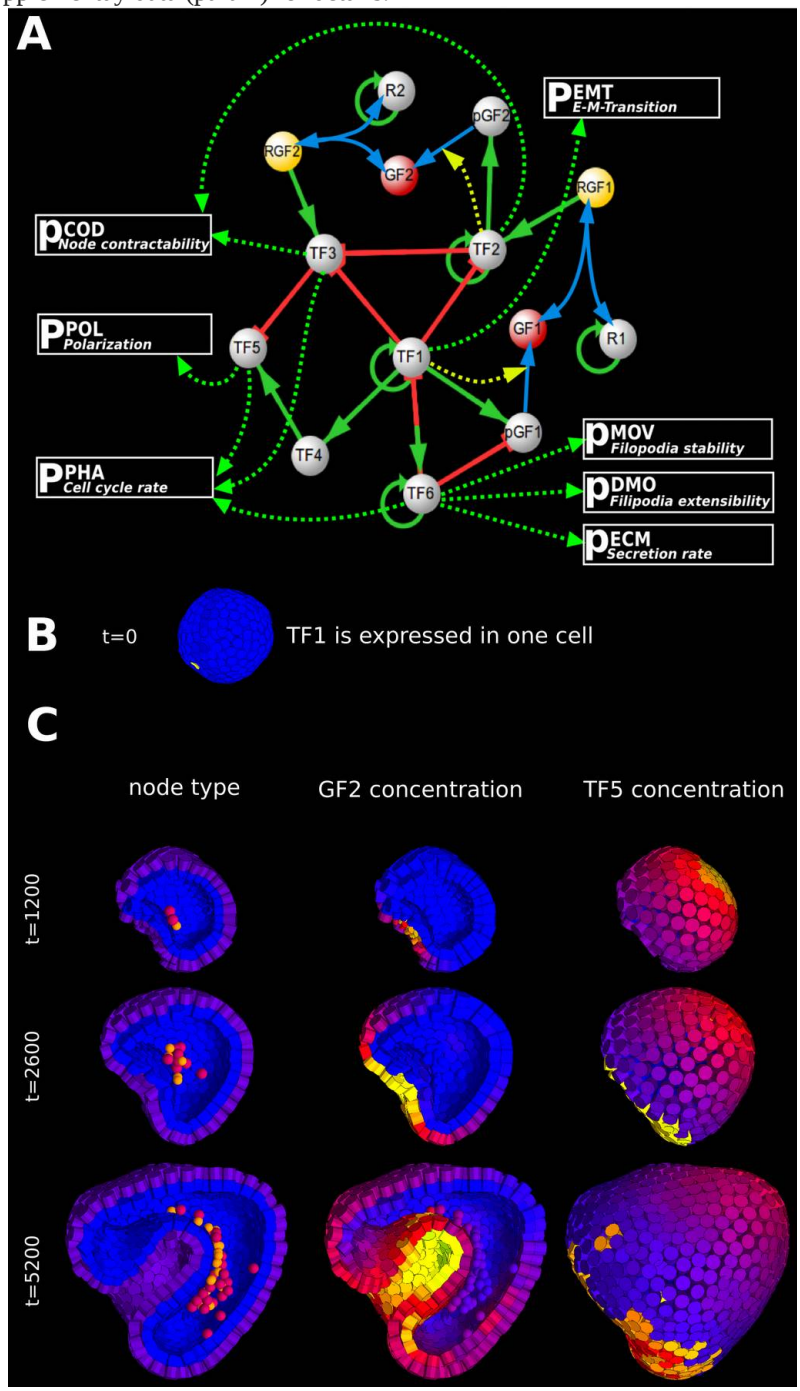
Supplementary Figure 15. Cell migration genotype phenotype map. Different simulations were run starting from the same initial conditions (left picture) except for one parameter that was modified. Each column shows the final phenotype when a single parameter is modified by a percentatge relative to the value of the wildtype (central row of pictures). A mass of three mesenchymal cells (green) migrates over a hexagonal epithelial sheet that has a gradient in the expression of its adhesion molecles. Thus, by means of noise and differential adhesion, mesenchymal cells get displaced, from upper right part of the epithelium, towards areas with high concentration of adhesion molecules (yellowish nodes). Notice that effective cell migration is only promoted by some parameter combinations (e.g. very short pseudopodia (p^{DMO} < wild type p^{DMO}) prevents migration).



Supplementary Figure 16. Genotype-phenotype map exploration for the differential adhesion mechanism. Different simulations, changing one parameter at a time, where run starting from the same initial conditions (left picture). Each column shows the final pattern when a single parameter is modified by a percentage relative to the value of the wild-type (central row of pictures). The parameters explored are, from left to right: heterotypic affinity between the two types of adhesion molecules, the maximum radius of interaction between nodes, the maximum length of movement by noise per time step, and the likelihood of the node to make an unfavorable movement by noise (see Text S1). The lack of final phenotypes corresponding to -60% and -90% decreases of maximum radius of interaction between nodes stems from the fact that in those cases, the radius of interaction is lower than the radius of equilibrium, thus preventing interactions between nodes (cells fail to keep their internal cohesion).



Supplementary Figure 17. An example developmental mechanisms with a gene network combining different cell behaviors. A) Schema of the gene regulatory network. TF, transcription factor, pGF, Growth factor transcript, GF, secreted growth factor, R, receptor, RGF, receptor-ligand complex. Solid green and red arrows depict positive and negative transcriptional regulation respectively. Solid blue arrows depict chemical reactions. Yellow dashed arrows indicate catalysis. Green dashed lines indicate regulation of cell behaviors or node properties. B) Initial conditions, single cell (yellow) expresses gene TF1. C) Outcome, after different number of iterations, of the developmental mechanism in A on the initial conditions in B. The left column shows, in section, the node types. Blue for basal side of cylinders, violet for the apical side of cylinders, red for mesenchymal cells and orange for extracellular matrix nodes. Middle and lower row display concentrations of GF2 and TF5 respectively (yellow for high concentration, blue for low concentration). See supplementay data (part B) for details.



Supplementary Figure 18. Simulation of a bead experiment in a "tooth-like" system. A growing epithelial bud is in contact with a growing mesenchymal condensation. In this simulation each cell is made of a single element. Mesenchymal cells express a transcriptional factor that activates its own transcription and the transcription of a growth factor precursor. This precursor requires a chemical modification to become a growth factor that diffuses between cells in the extra-cellular space. This reaction is catalyzed by the transcriptional factor. The growth factor diffuses in the extracellular space and binds to its receptor. This receptor is expressed only in the epithelium (where an auto-activatory transcriptional factor promotes its expression). The activated receptor enhances cell division in the epithelium. This leads to the bending of the epithelium close to the mesenchyme and finally enclosing the mesenchymal condensate (since there is an adhesion affinity between both tissues). Each row shows a time sequence. First row simulates development without bead addition, second to fifth show the same simulation but adding a bead at time step 10000 that releases the same growth factor the mesenchyme is secreting (beads with different concentrations of the growth factor). At the highest concentrations an ectopic epithelial bud can be seen beneath the bead. All pictures show a section of a three-dimensional system.