Kidney Morphogenesis Cellular Automaton - Model Details

Ben Lambert

May 16, 2014

1 Introduction

The aim of this model is to recapitulate the development of the Uretic Bud *in vivo* in 2D. However, it is also hoped that the model may be useful for the study of *in vitro* explants. The details of the model are described in brief below:

- A 2 dimensional rectangular domain, where the points in the array represent the presence of:

  - Epithelium cells - which consume GDNF

  - Metanephric mesenchymal cells (MM) - which produce GDNF

  - Extracellular matrix (ECM) - which allow for free diffusion of GDNF

- Epithelium cells comprising either:

  - A *flat* membrane at the base of the simulation domain (aimed at recapitulating the *in vivo* conditions from which the UB initially forms *in utero*.)

  - A mass of epithelium cells with in general, curved boundaries, suspended towards the centre of the simulation domain (representative of the *in vitro* explant experiment initial conditions.)

- A diffuse collection of mesenchymal cells initially separate from the epithelium.

2 GDNF field

Cells in the MM produce GDNF, and it diffuses freely across the ECM layer, and through the epithelium (being consumed by the latter). The reaction-diffusion type equation here will be assumed to be in equilibrium, since the diffusive timescale of GDNF is much less than that of the timescale of cell division. As such, the following is the form of the equation being solved:

$$D_G \nabla^2 = \Phi_G \tag{1}$$

Where in (1) D_G refers to the diffusion coefficient for GDNF, and Φ_G is the local rate of GDNF production or consumption (dependent on whether the cell in question is epithelium or mesenchyme). Specifically the rate of GDNF consumption

is assumed to have the following form:

$$\Phi_G = \begin{cases} K_G G, & \text{for epithelium.} \\ -\rho_G, & \text{for mesenchyme.} \\ 0, & \text{for the extracellular matrix.} \end{cases} \tag{2}$$

It is assumed that the rate of GDNF consumption is linearly-dependent on the concentration of substrate. This assumption is likely violated when the GDNF concentration is high, and the cell Ret-receptors are saturated. In later versions of this model, it may be better to assume a Hill-type reaction rate. (Although it is unclear as to what range of concentrations of GDNF are likely to saturate the Ret receptors, and whether these are likely to be encountered either *in vivo* or in experiments conducted thus far *in vitro*.) It is assumed to begin with that the mesenchyme GDNF production rate is constant, partly for simplicity, although this will be relaxed in latter models which, for example, allow mesenchymal GDNF production rates to be a positive function of their proximity to epithelium cells (perhaps as a proxy for Wnt11 levels).

The equation in (1) is non-dimensionalised using the following transformations:

$$\eta = \frac{x}{\Delta} \tag{3}$$

$$g = \frac{G}{G_x} \tag{4}$$

Where in (3) Δ refers to the typical cell dimensions (approximated as 5 μm),and G_x is the concentration of GDNF typically found *in vivo* (a value is currently not assigned here, since it is not strictly necessary for the non-dimensionalised simulation). These transformations result in the following non-dimensional form of the steady-state reaction-diffusion equation:

$$\nabla_\eta^2 g = \frac{\phi_g}{d_g} \tag{5}$$

Where in (5), $d_g = \frac{D_G}{K_G \Delta^2}$, $\phi_g = \frac{\Phi_G}{K_G G_x}$, and ∇_η^2 is the laplacian in the non-dimensional η spatial coordinates.

The boundary conditions which are assumed are: no-flux at the base of the Wolffian Duct, and periodic boundary conditions at either width. A finite difference approximation is used, with the no-flux boundary conditions at the top and bottom of the domain represented by the following relations for the non-dimensional GDNF field:

$$g_{0,j} = g_{1,j}$$
$$g_{M,j} = g_{M+1,j}$$

where this holds $\forall j = 1, ..., N$. M here is the depth of the simulation domain in terms of cell dimensions. Similarly, for the periodic boundary conditions:

$$g_{i,0} = g_{i,N}$$

$$g_{i,N+1} = g_{i,1}$$

where this holds $\forall i = 1, ..., M$. N here is the width of the simulation domain in terms of cell dimensions.

The finite-difference approximation for solving the steady-state reaction-diffusion equation is hence of the form:

$$g_{i+1,j} + g_{i-1,j} + g_{i,j+1} + g_{i,j-1} - (4 + \epsilon_{i,j})g_{i,j} = -\psi_{i,j} \tag{6}$$

Where in (6):

$$\epsilon_{i,j} = \begin{cases} \frac{1}{d_g}, & \text{for epithelium.} \\ 0, & \text{for mesenchyme.} \\ 0, & \text{for the extracellular matrix.} \end{cases} \tag{7}$$

Also, in (6):

$$\psi_{i,j} = \begin{cases} 0, & \text{for epithelium.} \\ \frac{\gamma}{d_g}, & \text{for mesenchyme.} \\ 0, & \text{for the extracellular matrix.} \end{cases} \tag{8}$$

Where in (8), $\gamma = \frac{\rho_G}{K_G G_x}$ captures the relative rate of GDNF production by the mesenchyme compared to the rate of GDNF consumption by epithelium under normal cellular conditions. It is assumed that $\gamma = 1$ in the simulations, in the absence of information regarding the relative weight of these two mechanisms. The diffusion rate constant, D_G, is assumed to be 10 $\mu m^2 s^{-1}$ (following Menshykau and Iber), and $K_G = 4 \times 10^{-2} s^{-1}$ (for this I currently have not thought of a way to estimate this parameter value - the *in vitro* experiments I have read have, thus far, not provided enough relevant information); making $d_g = 100$.

In the simulation, it is assumed that the GDNF field does not vary significantly within a particular time step. As such (in order to improve computational speed), the GDNF field is only updated at the end of each time step, not after each cell movement or proliferation. Ultimately however, it will be important to test whether the model conclusions are sensitive to this assumption.

3 Epithelium cell behaviour

3.1 Algorithm governing cellular behaviour

In this section I describe the behaviours of the epithelium cells in the cellular automaton model. In the simulation, each epithelium cell is visited in a random order, (actually epithelium and mesenchyme are updated together, so the randomised order corresponds to both epithelium and mesenchymal cells), and updated at each discrete time step according to the following pseudo-algorithm:

- *Available cells* - Are neighbouring cells available for movement or proliferation? If yes, proceed to step 2. If no, move on to next cell. Whether a cell is available depends on the specific rules in place (see section 3.2 for more details).

- *Move or proliferate* - Draw a random number $X \sim unif(0, 1)$. If $P_{move} > X$ then proceed to *move*. If not, proceed to *proliferate*.

- *Move*

  1. *Calculate the probability that a move takes place, P_m* - Based on the specific rules in place calculate the probability that a move takes place (see section 3.3 for more details).

  2. *Does a move take place?* - Draw a random number $X \sim unif(0, 1)$. If $P_m > X$ then proceed to next step. Otherwise, consider next epithelium cell.

  3. *Calculate probability of moving to each of the available cells* - Based on the specific rules being used, calculate a probabilities of moving to each of the available target cells (see section 3.4 for more details about the options for rules used here).

  4. *Choose amongst the available cells in accordance to their probability and move the cell in question.*

- *Proliferate*

  - The framework is exactly analogous to that of *Move* detailed above. However, the rules chosen for proliferation or moving are allowed to be different in the simulation. In proliferation, the mechanism is slightly different due to the fact that a daughter cell is created in the target cell, rather than a cell simply moved there.

- Consider next cell, returning to first step.

The program has been created in Matlab which allows for a number of permutations of the rules to be tried by the user. However, there are some combinations which are not allowed due to internal inconsistency amongst them. The idea is to allow the model's results to be easily tested for robustness to assumptions made.

3.2 Rules governing whether neighbouring cells are available

In this section I will detail the various different rules which govern into which cells an active cell (the one being updated) can move or proliferate. These rules are allowed to be different for cells which are moving or proliferating.

1. *Number of neighbours* - the choice here governs whether 4 (up, down, left, and right) or 8 (all points of a compass) are surveyed as potential candidates for a move or proliferation.

2. *Connectivity of cells*

   (a) All vacant cells are allowed.

   (b) Moves or proliferations are allowed into vacant cells only if the active cell (the one being moved, or the daughter cell being created) are connected to other cells. Here connectivity means that there is at least one 4-neighbour (up, down, left or right) which is occupied by an epithelium cell

   (c) Same as above, but more stringent. Only allow moves if all cells are connected. This allows for the possiblity of a move leaving one cell unconnected.

   (d) All vacant cells and mesenchyme are allowed as possible move/proliferation locations. If a movement into a cell occupied by a mesenchymal cell occurs, then the mesenchyme cell is removed from the simulation.

   (e) Same as 2d, although only allow movement into a space if that active cell remains connected.

   (f) Same as 2e, although instead of killing mesenchyme cell (upon movement into its space), move the mesenchyme into a vacant space neighbouring it. If there are no vacant spaces available for the mesenchyme cell to move into, then the move is not allowed. If multiple spaces are available for the cell to move into, then choose one of them at random.

The above rules 2b, 2c, 2e and 2f are phenomenlogical, but are aimed at mimicking the cell-cell adhesion experienced by the epithelium. Rules 2d, 2e, 2f are included as to allow an interaction between the epithelium cells and the mesenchyme, when the former move into the area occupied by the latter. It is thought that 2f is the most biologically realistic, since excess cell death is not typically observed. However, the current way in which cells are chosen for the moved mesenchyme is thought to be an over-simplification, since it is likely that cells which are displaced move more frequently in a direction in which they are pushed. Later models may give more weight towards those directions probabilistically.

3.3 Rules governing the calculation of P_m or P_p - the probability of a move or proliferation occurring after the action has been chosen

In this section I describe the available choices of model rules which govern the calculation of the probability of a move/proliferation occurring takes place. To be clear, this is the step after the choice has been made to either go down the algorithmic branches corresponding to *Move* or *Proliferate* respectively. It is chosen to make the decision as to whether to continue along each branch here, rather than at the stage before to allow for potentially different rules governing moving and proliferating respectively. The rules available for calculating whether or not a move/daughter cell creation takes place are given below:

1. The probability of a move/proliferation is a constant specified by the user.

2. The probability of a move/proliferation is a positive function of the local GDNF concentration (the level of GDNF for that cell at that grid point). The equation used here to specify the probability is that of a probit model

$$P(action) = \Phi(c_1 + c_2 g) \tag{9}$$

Where in (9), Φ is the standard normal CDF, and *action* can either be a move or a proliferation event.

3. Move/proliferation probability is determined by the sum of all local positive GDNF gradients, between each of the available neighbours and the current location. Again, the model used here for the probability is the probit model:

$$P(action) = \Phi(c_1 + c_2 \sum_{allowed} (g^{neighbours} - g^{current}) \times \mathbb{1}(g^{neighbours} > g^{current}) \tag{10}$$

Where in (10), the indicator function $\mathbb{1}(g^{neighbours} > g^{current})$, is equal to one if the gradient in GDNF allowed by the move is positive, and zero otherwise.

Rule 2 is aimed at allowing more cellular activity (moves or proliferations) in areas where GDNF concentration is higher. The use of a normal CDF here is partly to constrain the probability to lie in the appropriate (0,1) range, but can also act to provide thresholding-type behaviour, which may be necessary to generate discrete bud formation from the Wolffian Duct *in vivo* opposed to a general 'swelling'. Rule 3 allows more cellular activity in areas where there is more GDNF to be gained by a potential move or proliferation. I am less sure as to the biological realism of this latter rule, although for it could be thought of as a type of chemotaxis for moving cells, or orientated cell division for proliferating cells.

3.4 Rules governing selection of move target or daughter cell location

This last section on rules details how to determine the location (amongst the available, allowed neighbouring cells) into which either a cell should move or create a daughter cell. The rules governing how the probabilities of choosing a given target location are given below:

1. The probabilities of choosing each of the possible target cell are equal, and given by:

$$P(target) = \frac{1}{N_{target}} \tag{11}$$

Where in (11), N_{target} specifies the total number of allowed possible target cells.

2. The weights given to a specific target cell are given by:

$$W(target) = \Phi(c_3 + c_4(g^{target} - g^{current})) \tag{12}$$

The weights across all targets are then used as weights in a Dirichlet distribution to calculate probabilities which sum to 1; with the higher weights corresponding to higher probabilities.

3. The weights given to a specific target cell are related to the percentage changes in GDNF:

$$W(target) = \Phi(c_3 + c_4 \frac{g^{target} - g^{current}}{g^{current}}) \tag{13}$$

The weights across all targets are then used as weights in a Dirichlet distribution to calculate probabilities which sum to 1; with the higher weights corresponding to higher probabilities.

4. The probabilities of choosing a specific target cell are given by a multinomial logistic distribution:

$$P(target_n) = \frac{exp(c_3 + c_4(g^{target} - g^{current}))}{\sum\limits_{alltargets} exp(c_3 + c_4(g^{target} - g^{current}))} \tag{14}$$

Where in (14), the benefit of this particular form is that the probabilities naturally sum to 1.

Rules 2, 3, and 4 represent chemotaxis (when considering movement), and orientated cell division (when considering proliferation). It is not clear to me whether orientated cell division can occur along concentration gradients, but this is provided for model completeness.

After conducting the simulations, it does not appear that there are significant differences between these different methods. In terms of those rules which weight towards those target cells which have the highest (positive) gradient in GDNF, the preferred rule is 4, which naturally provides a way of creating probabilities which sum to 1 overall.

4   Mesenchyme cell behaviour

The current version of the simulation does not allow for 'non-directed' behaviour of the mesenchyme. Their only movement behaviour is their ability to be either 'killed' or moved out of the way, by advancing epithelium cells. Later versions of the simulation will allow for autonomous mesenchymal behaviour; for example attraction towards epithelium (or Wnt11 released by the epithelium), or other mesenchyme cells.

5   Cell death

The current version of the simulation does not allow cell death directly; only indirectly through the epithelium's ability to (dependent on the exact rules implemented) 'overtake' a position currently occupied by the mesenchyme. Later versions will allow the user to stipulate differing death rates for the epithelium and mesenchyme, and also allow these to be context-dependent (dependent on, for example, the connectivity of the cell in question).

6   Creation of the epithelium layer

Two different schema are considered for the creation of an epithelium layer:

- A flat layer of epithelium at the base of the simulation domain to emulate *in vivo* formation of the initial Uretic Bud. Although this layer is *in rerum natura* mono-layered, I choose to use a multi-layered epithelium. This is to allow for the fact that the simulation domain is two dimensional whereas in practice it is three dimensional. Making the layer multi-layered allows cells which are outside of the simulation plane to pass into the layer bordering on the ECM/mesenchyme.

- A randomised mass of epithelium cells towards middle of the simulation domain representing *in vitro* experimental conditions. The idea here is that branching arises partly as a result of local depletion of GDNF concentration in areas 'crypts' of epithelium next to outgrowths.