Supplementary material

This document provides supplementary information on the dynamic modelling of non-photoconverted and photoconverted (green/red) neutrophils for describing cell migration, using the drift-diffusion equation.

Model formulations

We make the assumption here that the neutrophil dynamics can be described by a random walk process. Hence, the models of cell migration are derived from the drift-diffusion equation,

$$\frac{\partial p(x,t)}{\partial t} = b\frac{\partial^2 p}{\partial x^2} + a\frac{\partial p}{\partial x}, \tag{1}$$

where $p(x,t)$ is the proportion of cells found at position x at time t, b is the effective diffusion coefficient of the cells and a is rate of drift (as would be due to any chemotactic bias). We solve this equation in two cases, for green (non-photoconverted) and red (photoconverted) cells, under separate conditions to yield models for the population distribution of the neutrophils changing over time. From the resulting expressions we estimate the drift and diffusion parameters using least-squares methods.

Model of green cell migration

If we assume that the green cells have an arbitrary distribution, which is far enough away from the boundary that some time elapses before any cell reaches the boundary, then using the method described in [1] we can multiply Eqn. (1) by x and integrate to yield,

$$E\left[x(t)\right] = at + c_1, \tag{2}$$

where the parameter a characterises the drift of the neutrophil population (see Eqn. 1) and c_1 is an offset term. Eqn. (2) implies that the mean position of cells will be constant if there is no bias - the drift rate parameter $a = 0$ (diffusion only), but will move linearly with time in the direction of any bias for the case when a is non-zero.

The elements of the set of cell positions $\mathcal{X}(t) = \{x_1(t), \ldots, x_{n_t}(t)\}$ (where n_t is the number of observed cells at time t) are averaged to produce an observed value of mean cell distance from the wound $\bar{x}(t)$ for each t. This is the data derived version of the theoretical quantity $E\left[x(t)\right]$ found in the model equations. If $\bar{x}(t)$ is plotted against time (see Figs. in main article) it decreases linearly up to a certain switching time, say $t = t_S$. A suitable value of t_S was determined by visual inspection of the graph in each case.

In order to identify the bias parameter, a, that characterises the drift of the neutrophil population, we used weighted least squares estimation (see for example [3]), where we define,

$$\mathbf{z} = \begin{pmatrix} \bar{x}(t_1) & \ldots & \bar{x}(t_S) \end{pmatrix}^T, \tag{3}$$

$$\Phi = \begin{pmatrix} t_1 & t_2 & \ldots & t_S \\ 1 & 1 & \ldots & 1 \end{pmatrix}^T, \tag{4}$$

$$\theta = \begin{pmatrix} a & c_1 \end{pmatrix}^T. \tag{5}$$

It should be noted here that in keeping with the experimental observation rate one time unit $(t_k - t_{k-1})$ is 300 seconds. Then model (2) implies that

$$\mathbf{z} = \Phi\theta + \mathbf{e}, \tag{6}$$

where $\mathbf{e}$ represents model error. In the case of the green cells, the variability in the number of visible cells was due to cells entering and leaving the domain as well as cells

becoming indistinguishable at time from one another. We therefore defined a weighting vector related to deviation from the mean number of visible cells, as follows:

$$W = \text{diag}([w_1 \dots w_S]), \tag{7}$$

$$w_k = 1 - \left(\frac{\bar{n} - n_k}{\bar{n}} \right)^2, \tag{8}$$

$$\bar{n} = \frac{1}{S} \sum_{k=1}^{S} n_k, \tag{9}$$

where diag(.) is the diagonal matrix constructed from its vector argument and n_k is the number of cells visible at time $t = t_k$. Then the best estimate $\hat{\theta}$ of θ is given by,

$$\hat{\theta} = (\Phi^T W \Phi)^{-1} \Phi^T W \mathbf{z}. \tag{10}$$

To assess confidence in the estimated parameters we calculated the standard deviation σ from the residuals using,

$$e(j) = \bar{x}(t_j) - (\hat{a}t_j + \hat{c}_1), \tag{11}$$

$$\mathbf{e} = (e(1) \dots e(S))^T, \tag{12}$$

$$\sigma = \sqrt{\frac{\mathbf{e}^T \mathbf{e}}{S - m} (\Phi^T \Phi)^{-1} \Phi^T W W \Phi (\Phi^T \Phi)^{-1}}, \tag{13}$$

where m is the number of parameters in the model. In this case $m = 2$.

Model of red cell migration

If we assume that the wound in the fish can be mathematically characterised as a reflecting boundary and that the initial distribution of cells is a delta distribution at the origin, then solving Eqn. (1) using the method of reflections [2] we can describe the evolution of the distribution of red cells as

$$p(x,t) = \frac{1}{\sqrt{4\pi bt}} \left(e^{\frac{-(x-at)^2}{4bt}} + e^{\frac{-(x+at)^2}{4bt}} \right). \tag{14}$$

Multiplying Eqn. (14) by x^2 and then integrating over $x \in [0, \infty)$ yields,

$$E\left[x(t)^2\right] = a^2 t^2 + 2bt + c_2. \tag{15}$$

Eqn. (15) describes $E[x(t)]$ as a quadratic form if $a > 0$ (bias present away from the wound) and a linear form if $a = 0$ (no bias / diffusion only).

For the purposes of linear estimation, in the $a = 0$ case we define the regression matrix as,

$$\Phi = \begin{pmatrix} t_1 & t_2 & \dots & t_N \\ 1 & 1 & \dots & 1 \end{pmatrix}^T. \tag{16}$$

If a is non zero we have

$$\Phi = \begin{pmatrix} t_1^2 & t_2^2 & \dots & t_N^2 \\ t_1 & t_2 & \dots & t_N \\ 1 & 1 & \dots & 1 \end{pmatrix}^T. \tag{17}$$

Now, $z = \begin{pmatrix} \bar{x}(1)^2 & \dots & \bar{x}(N)^2 \end{pmatrix}^T$ and in this case we define $W = \text{diag}([n_1 \dots n_S])$ (this is because variability in the number of visible red cells is only down cells becoming obscured by proximity to each other. Very few if any leave the domain and no new cells

can enter. We can proceed analogously to the last section to find $\hat{\theta}_2 = \begin{pmatrix} 2\hat{b} & \hat{k} \end{pmatrix}^T$ or $\hat{\theta}_3 = \begin{pmatrix} \hat{a}^2 & 2\hat{b} & \hat{k} \end{pmatrix}^T$ and the associated confidence interval.

A higher order model will always give a better fit to a data set. It will have a smaller associated error function $J(m) = \mathbf{e}^T\mathbf{e}$ (where m is the order of the model). We used an F-test to evaluate the significance of the improvement [4]. The higher order model is better with $1-\alpha$ certainty if $f > \chi^2_\alpha(1)$, where $\chi^2_\cdot(.)$ is the Chi-Squared distribution and

$$f = \frac{J(m-1) - J(m)}{J(m)}(N-m). \tag{18}$$

Typically, if an F-test returns a value $f > 4$ then the test implies that the more complex model should be accepted - in this case that the quadratic model should be chosen over the linear model [4].

References

[1] Codling EA, Plank MJ, Benhamou S: **Random walk models in biology**. *Journal Royal Society Interface* 2008, **5**:813–834.

[2] Coleman MP: *An introduction to partial differential equations with MATLAB*. CRC Press 2005.

[3] Nelles O: *Nonlinear System Identification*. Springer 2001.

[4] Soderstrom T, Stoica P: *System Identification*. Prentice Hall International 1989.