S1 Supplementary model information

This supplementary document describes the model for neural crest cell migration with discrete leader-follower states, as used in (McLennan *et al.* 2015a,b, 2017). For the continuous state models, everything applies as stated here, apart from the choice of direction of movement, which is described in the main text.

S1.1 Model equations

Equations for tissue growth and reaction-diffusion of the chemoattractant were used as previously described (McLennan *et al.* 2015a,b, 2017, Supplementary Information), and are reproduced below.

S1.1.1 Domain growth

Tissue growth was modelled as uniform, with the length of the migratory domain at any time between $t = 0$ and $t = 24$ hours given by the logistic equation

$$L_x(t) = L_0 \left(\frac{L_\infty e^{a(t-t_s)L_\infty}}{L_\infty - 1 + e^{a(t-t_s)L_\infty}} + 1 - \frac{L_\infty e^{a(-t_s)L_\infty}}{L_\infty - 1 + e^{a(-t_s)L_\infty}} \right), \tag{1}$$

with parameters $L_0 = 300\mu\text{m}, a = 0.08\text{h}^{-1}\mu\text{m}^{-1}, t_s = 16$ hours, $L_\infty = 870$, determined by least-squares fitting to experimental domain length measurements (McLennan *et al.* 2012).

S1.1.2 Chemoattractant reaction-diffusion

To model the change in chemoattractant concentration on a growing domain with $(x, y) \in [0, L_x(t)] \times [0, L_y]$, we rescale the growing domain to a stationary domain of unit length in x. To maintain numerical accuracy as the effective lattice spacing increases due to the rescaling, we use a solver with automatic grid refinement (d03ra from the Numerical Algorithms Group (NAG)), as was done is studies using the previous model (McLennan *et al.* 2012). Omitting the explicit time dependence of L_x, the change of chemoattractant concentration at a point $(x, y) \in [0, 1] \times [0, L_y]$ of the stationary domain is given by the RDE

$$\begin{aligned} \frac{\partial c}{\partial t} = D \left(\frac{1}{L_x^2} \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) - c \sum_{i=1}^{N(t)} \frac{\lambda}{2\pi R^2} \exp \left[-\frac{L_x^2 (x - x_i)^2 + (y - y_i)^2}{2R^2} \right] \\ + \chi c(1 - c) - \frac{\dot{L_x}}{L_x} c, \end{aligned} \tag{2}$$

where the terms on the right-hand side describe diffusion, internalisation, production and dilution (by tissue growth, the dot denoting time derivative), respectively. Scaling factors of L_x are introduced by rescaling to a stationary domain to solve numerically on a grid of unit length (McLennan *et al.* 2012). Parameter names and values are given in Table 1.

S1.2 Sensing accuracy

Berg & Purcell (1977) derive a fundamental biophysical limit to the accuracy with which a cell can sense a chemical gradient. We briefly outline their derivation here before commenting on parameterisation. For a more detailed derivation, see the original work (Berg & Purcell 1977).

The fundamental limit in the accuracy of concentration measurements is due to fluctuations in the numbers of molecules measured. Fluctuations in particle number, N, are proportional to $\sqrt{N}$. To proceed with the derivation, consider a sensor counting N molecules in a volume V with background (or average) concentration $\bar{c}$. The inaccuracy in a single concentration measurement is

$$\frac{\Delta c}{\bar{c}} \approx \frac{1}{\sqrt{N}} = \frac{1}{\sqrt{V\bar{c}}}, \tag{3}$$

in three dimensions, or $1/\sqrt{A\bar{c}}$ in two dimensions, where A is the measurement area. The count of molecules can be improved by repeated measurements. A sensor counting molecules in a volume can make $n = TD/V^{2/3}$ independent measurements in a time T, based on the timescale of a molecule diffusing through the measurement volume V. This improves the (root mean square) measurement error by $1/\sqrt{n}$ (Berg & Purcell 1977). Thus, with $V \sim R^3$, the measurement uncertainty reduces to

$$\frac{\Delta c}{\bar{c}} \approx \frac{1}{\sqrt{DT\bar{c}R}} =: \zeta, \tag{4}$$

in three dimensions, or $\Delta c/\bar{c} \approx 1/\sqrt{DT\bar{c}}$ in two dimensions. Here we have introduced the dimensionless parameter ζ, which depends on the background concentration, c. To account for the dynamically changing background concentration, we explicitly scale ζ in simulations by the current relative concentration, i.e., $\sqrt{c_0/c}$, where c_0 is the starting background concentration (see Section S1.4 for pseudocode). The exact derivation of the sensing accuracy introduces a numerical factor of order unity, but since we can only parameterise the sensing accuracy to orders of magnitude (see Section S1.5.2), we ignore this.

S1.3 Integrate & switch mechanism

In McLennan *et al.* (2015b) a variable that records for how long each cell has been exposed to the presence of a chemoattractant. This variable increases at a fixed rate when a chemoattractant gradient above the sensing accuracy threshold is sensed, and decreases otherwise at another fixed rate. These rates are inversely proportional to the parameters 'leader-to-follower switching time', τ_{LF}, and 'follower-to-leader switching time', τ_{FL}, respectively. Thus, this variable effectively integrates the time spent in a chemoattractant gradient (with a decaying "memory"), though this could be easily modified to instead record the magnitude of the gradient or absolute value of the concentration.

Once the net time spent in a chemoattractant gradient (with time in the absence of an increasing gradient counting negatively) reaches a threshold, follower cells switch state to adopt leader behaviour, i.e., begin to undergo chemotaxis. Once cells are in a leader state and remain in a positive chemotaxis gradient, they do not increase their signal sensed further, that is, they do not become further entrained to stay in leader state, which would increase the time taken to switch back to a follower state in the absence of the gradient. Similarly, the intracellular signal decays with time spent in the absence of a positive gradient, until it becomes low enough for cells to switch to a follower state, and then does not decrease further (unless a gradient is found again). As a consequence, cells that have just switched state cannot switch back immediately, as long as the directional signal is lost/gained on timescales shorter than the switching time.

S1.4 Pseudocode

For model simulations, this code was implemented in Mathwork's MATLAB, and the chemoattractant profile was solved using the Numerical Algorithms Group's (NAG) d03ra, as previously described (McLennan *et al.* 2012, 2015a,b). In the interest of reproducibility, rather than just repeatability, we give the pseudocode to be implemented in the reader's programming language of choice.

main function

main function

initialise model parameters and first cells ▷ see Table 1
for $t = 6$ to 24 **do**
    if t = insertion time **then**
        if there is space to insert a cell **then**
            insert a new cell at start of domain
        end if
    end if
    solve chemoattractant profile ▷ see Eq. (2)
    grow domain, update cell positions ▷ see Eq. (1)
    move cells
    integrate-and-switch
end for

move cells

for $i = 1$ to number of cells **do**
    pick a cell at random without replacement
    pick n_{filo} random directions
    if cell is a leader **then**
        measure chemoattractant concentration at cell position,

$$c_{\mathrm{old}} = \int c(x,y) \exp\left[-\frac{x^2+y^2}{2R^2}\right] \mathrm{d}x\,\mathrm{d}y$$

        measure chemoattractant concentration in random direction(s) at distance l_{filo} away (pick highest), c_{new}

if $\frac{c_{\text{new}}-c_{\text{old}}}{c_{\text{old}}} \geq \zeta\sqrt{\frac{c_0}{c_{\text{old}}}}$ **then** ▷ gradient is above (scaled) sensing accuracy
move in chosen direction
else ▷ cell has not found a favourable chemoattractant gradient
move in random direction
end if
else if the cell is attached **then** ▷ cell is following another cell
if other cell is within l_{filo} **then**
move in same direction as other cell
else ▷ other cell is out of reach
detach cell
end if
else ▷ the cell is a detached follower
check if there is another cell in random direction(s) at distance l_{filo} (pick closest)
if a cell was found and is a leader (or part of a chain in contact with a leader) **then**
move in same direction as other cell
else
move in a random direction
end if
end if
end for

integrate-and-switch

increase signal sensed for cells that have sensed a chemoattractant gradient (but not beyond the upper threshold)
decrease signal sensed for cells that have not sensed a chemoattractant gradient (but not below the lower threshold)
followers whose signal sensed is at the upper threshold switch to become leaders
leaders whose signal sensed is at the lower threshold switch to become followers

Note on volume exclusion: Any attempted movement is aborted if it would lead to overlap with another cell or the domain boundary.

S1.5 Parameterisation

See Table 1 for values of parameters used in model simulations, and below for explanations of selected parameters.

Table 1: **Model parameters**
Parameter values listed were used as a default, unless otherwise stated. Where a range is given, the model gives qualitatively similar results within that range, and the value in parentheses is the one used as a default. The chemoattractant concentration is implemented in relative units, such that the starting value $c_0 = 1$.

	Description	Value	Reference
n_{filo}	directions sampled per time-step	2	n/a, Section S1.5.1
Δt	simulation time-step	1 min	n/a
R	cell radius (nuclear)	7.5 μm	McLennan & Kulesa (2010)
v_{lead}	cell speed (leader cells)	41.6 μm/h	**?**
v_{follow}	cell speed (follower cells)	49.9 μm/h	**?**
L_y	height of domain	120 μm	McLennan *et al.* (2012)
L_x	length of domain (grows, Eq. (1))	300 to 1100 μm	McLennan *et al.* (2012)
l_{filo}	sensing radius	27.5 μm	Section S1.5.1
$l_{\text{filo}}^{\max}$	max. separation of cells in contact	45μm	Section S1.5.1
D	diffusivity of chemoattractant	0.1 to $10^5 \mu\text{m}^2$/h (0.1)	Section S1.5.1
χ	production rate of chemoattractant	10^{-8} to 1/h (10^{-4})	Section S1.5.1
λ	chemoattractant internalisation rate	10^2 to 10^4 (10^3) μm^2/h	Section S1.5.1
k_{in}	rate at which cells enter the domain	10/h	Section S1.5.1
ζ	sensing accuracy	0.001 to 0.25 (0.1)	Section S1.2

S1.5.1 Notes and further references

Experimental time Cell migration is assumed to start approximately six hours after electroporation ($t = 0$).

Directions sampled per time-step, n_{filo} This cannot be directly related to the number of filopodia, which are greater in number, but sample at a lower speed (McLennan *et al.* 2012).

Sensing radius, l_{filo} This was calculated as the sum of the cell radius (7.5 μm) and the mean filopodial length (which was directly measured from the cell body to be 9 μm and estimated from total cell size to be circa 20 μm). Since we have only implemented contact between filopodium and cell body, but not between two filopodia, which does occur *in vivo* (Teddy & Kulesa 2004), we allow for a greater effective length.

Maximum cell separation before contact is lost, $l_{\text{filo}}^{\max}$ The maximum cell size including filopodia was measured to be 86.3 μm, half of which gives an estimate of maximum cell separation of 43.15 μm. Independent measurements of filopodial lengths gave a maximum of 30.4 μm (from the cell body), which, together with the cell radius $R = 7.5\mu$m and the average filopodial length (allowing for interfilopodial contact) of 9 μm, gives an estimate of 46.5 μm.

Diffusion coefficient of chemoattractant, D The primary identified chemoattractant in chick cranial neural crest migration is VEGF^{165} (McLennan *et al.* 2010). Its related isoform VEGF^{164} is known to bind to ECM, and studies in angiogenesis estimate as little as 1% may be freely diffusing, the rest bound to ECM and cellular receptors (Mac Gabhann *et al.* 2006). Hence, we choose a low effective diffusivity. For freely diffusing VEGF *in vivo*, angiogenesis modelling studies have used much higher values of $10^5\,\mu\text{m}^2/\text{h}$ (Mac Gabhann *et al.* 2006, Jain & Jackson 2013). However, with diffusivities of that order of magnitude our model simulations still give qualitatively similar results.

Production rate of chemottractant, χ In other tissues, VEGF production, or estimates thereof, range from 0.01-0.20 molecules/cell/s (Yen *et al.* 2011), or 4.39-5.27$\cdot 10^{-5}$ molecules/μm$^-$2/s (Vempati *et al.* 2011) to $0.25 \cdot 10^{-17}$ pmol/μm^2/s(Mac Gabhann *et al.* 2006). In our system, the rate of VEGF production is unknown and difficult to measure. However, it is outweighed by internalisation through migrating neural crest cells, as VEGF is not seen to be replenished in trailing portions of the stream (McLennan *et al.* 2010). Thus, we assume χ to be low.

Chemoattractant internalisation rate, λ To our knowledge, no estimates or measurements of VEGF internalisation rate of chick cranial neural crest exist. Angiogenesis studies have used values of $k_{\text{VEGFR2}} = O(10^{-4})$/s per receptor (Mac Gabhann & Popel 2005, Yen *et al.* 2011). Berg & Purcell (1977) estimate the number of receptors needed for a near-optimal sensing accuracy as $N_R = R/s$, where R is the cell radius and s the receptor size. With $s = O(\text{nm})$, we can estimate the near-optimal number of receptors to be $N_R \geq 10^4$. If receptor internalisation rates are comparable to other tissues, a lower bound for the total internalisation rate would be given by $k_{\text{VEGFR2}} N_R \geq 1/\text{s}$ (per cell). From this we estimate the chemoattractant consumption, as defined in (2), to be $\lambda \geq O(10^3)\mu\text{m}^2/\text{h}$. However, the concentration of VEGF in our system is unknown, and hence the units of c, and therefore λ, in our model are arbitrary. We assume a high λ to ensure quick consumption of chemoattractant by cells.

Rate at which cells enter the domain, k_{in} This is the rate of attempted cell insertions, and not the effective rate seen *in vivo*. In a typical simulation, on the order of 10% of insertions are unsuccessful. Greater values of the insertion rate thus result in equal or only slightly increased cell numbers. It should be noted here again that our simulations are a two-dimensional abstraction of the three-dimensional migratory stream, which may contain 4-5 times as many cells *in vivo* in the transverse (z) direction. Thus, cell numbers in (unperturbed) simulations are approximately correct for a section of the migratory stream, to within the accuracy that total cell numbers are known *in vivo*.

S1.5.2 Parameterisation of the sensing accuracy

Most of the variables upon which the sensing accuracy depends are underdetermined in the case of chick cranial neural crest migration, such as VEGF diffusivity, D, VEGF background concentration, $\bar{c}$, and the sensing time, T. Nevertheless, we can proceed to estimate order of magnitudes, which can serve as bounds for our model simulations.

Background concentration, $\bar{C}$ The concentration of VEGF used in *in vitro* experiments is 1 μg/ml (McLennan *et al.* 2010), which, at a molecular weight of 19.2 kDa $\approx$ 20kg/mol, leads us to estimate $\bar{c} \approx 3 \cdot 10^7/\mu\text{m}^3$ (50mM).

Sensing time, T The time-step of our simulations is $\Delta t = 1$ minute, and we assume that a cell takes up only a fraction of this time with sensing, and most of it with movement. We could therefore estimate $T \leq 0.1 \cdot \Delta t = 0.1$ minutes. If we relax our assumptions, this estimate might change by an order of magnitude. This would only change the sensing accuracy by a factor of roughly 1/3, which gives qualitatively similar results in typical model simulations.

Lower bounds on gradient measurement accuracy For the measurement of a gradient, i.e., the difference between two concentration measurements, the Berg-Purcell limit (4) increases by a factor of $\sqrt{2}$.With the estimates for $\bar{c}$ and T as above, and the parameter values $D = 0.1\,\mu\text{m}^2/\text{h}$ and $R = 7.5\,\mu\text{m}$ (Table 1), we obtain an estimate of the sensing accuracy (4) of $\zeta_{d=3} \approx 0.002$ in three dimensions, or $\zeta_{d=2} \approx 0.01$ in two dimensions. These can be taken as a *lower bound* for the (order of magnitude of) sensing accuracy of neural crest cells in our model. Note that the sensing accuracy rescales with changing background concentration, which has to be taken care of in the computational implementation (see Section S1.4).

References

Berg, H. C. & Purcell, E. M. 1977. Physics of chemoreception. *Biophysical Journal*, **20**, 193–219. doi:10.1016/S0006-3495(77)85544-6.

Jain, H. V. & Jackson, T. L. 2013. A hybrid model of the role of VEGF binding in endothelial cell migration and capillary formation. *Frontiers in Oncology*, **3**, 102. doi:10.3389/fonc.2013.00102.

Mac Gabhann, F., Ji, J. W. & Popel, A. S. 2006. Computational model of vascular endothelial growth factor spatial distribution in muscle and pro-angiogenic cell therapy. *PLOS Computational Biology*, **2**, e127. doi:10.1371/journal.pcbi.0020127.

Mac Gabhann, F. & Popel, A. S. 2005. Differential binding of VEGF isoforms to VEGF receptor 2 in the presence of neuropilin-1: a computational model. *American Journal of Physiology. Heart and Circulatory Physiology*, **288**, H2851–60. doi:10.1152/ajpheart.01218.2004.

McLennan, R., Bailey, C. M., Schumacher, L. J., Teddy, J. M., Morrison, J. A., Kasemeier-Kulesa, J. C., Wolfe, L. A., Gogol, M. M., Baker, R. E., Maini, P. K. & Kulesa, P. M. 2017. DAN (NBL1) promotes collective neural crest migration by restraining uncontrolled invasion. *The Journal of Cell Biology*, **216**, 3339–3354. doi:10.1083/jcb.201612169.

McLennan, R., Dyson, L., Prather, K. W., Morrison, J. A., Baker, R. E., Maini, P. K. & Kulesa, P. M. 2012. Multiscale mechanisms of cell migration during development: theory and experiment. *Development*, **139**, 2935–44. doi:10.1242/dev.081471.

McLennan, R. & Kulesa, P. M. 2010. Neuropilin-1 interacts with the second branchial arch microenvironment to mediate chick neural crest cell dynamics. *Developmental Dynamics*, **239**, 1664–73. doi:10.1002/dvdy.22303.

McLennan, R., Schumacher, L. J., Morrison, J. A., Teddy, J. M., Ridenour, D. A., Box, A. C., Semerad, C. L., Li, H., McDowell, W., Kay, D., Maini, P. K., Baker, R. E. & Kulesa, P. M. 2015a. Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front. *Development*, **142**, 2014–2025. doi:10.1242/dev.117507.

McLennan, R., Schumacher, L. J., Morrison, J. A., Teddy, J. M., Ridenour, D. A., Box, A. C., Semerad, C. L., Li, H., McDowell, W., Kay, D., Maini, P. K., Baker, R. E. & Kulesa, P. M. 2015b. VEGF signals induce trailblazer cell identity that drives neural crest migration. *Developmental Biology*, **407**, 12–25. doi:10.1016/j.ydbio.2015.08.011.

McLennan, R., Teddy, J. M., Kasemeier-Kulesa, J. C., Romine, M. H. & Kulesa, P. M. 2010. Vascular endothelial growth factor (VEGF) regulates cranial neural crest migration in vivo. *Developmental Biology*, **339**, 114–125. doi:10.1016/j.ydbio.2009.12.022.

Teddy, J. M. & Kulesa, P. M. 2004. In vivo evidence for short-and long-range cell communication in cranial neural crest cells. *Development*, **131**, 6141–51. doi:10.1242/dev.01534.

Vempati, P., Popel, A. S. & Mac Gabhann, F. 2011. Formation of VEGF isoform-specific spatial distributions governing angiogenesis: computational analysis. *BMC Systems Biology*, **5**, 59. doi:10.1186/1752-0509-5-59.

Yen, P., Finley, S. D., Engel-Stefanini, M. O. & Popel, A. S. 2011. A two-compartment model of VEGF distribution in the mouse. *PLOS ONE*, **6**, e27514. doi:10.1371/journal.pone.0027514.