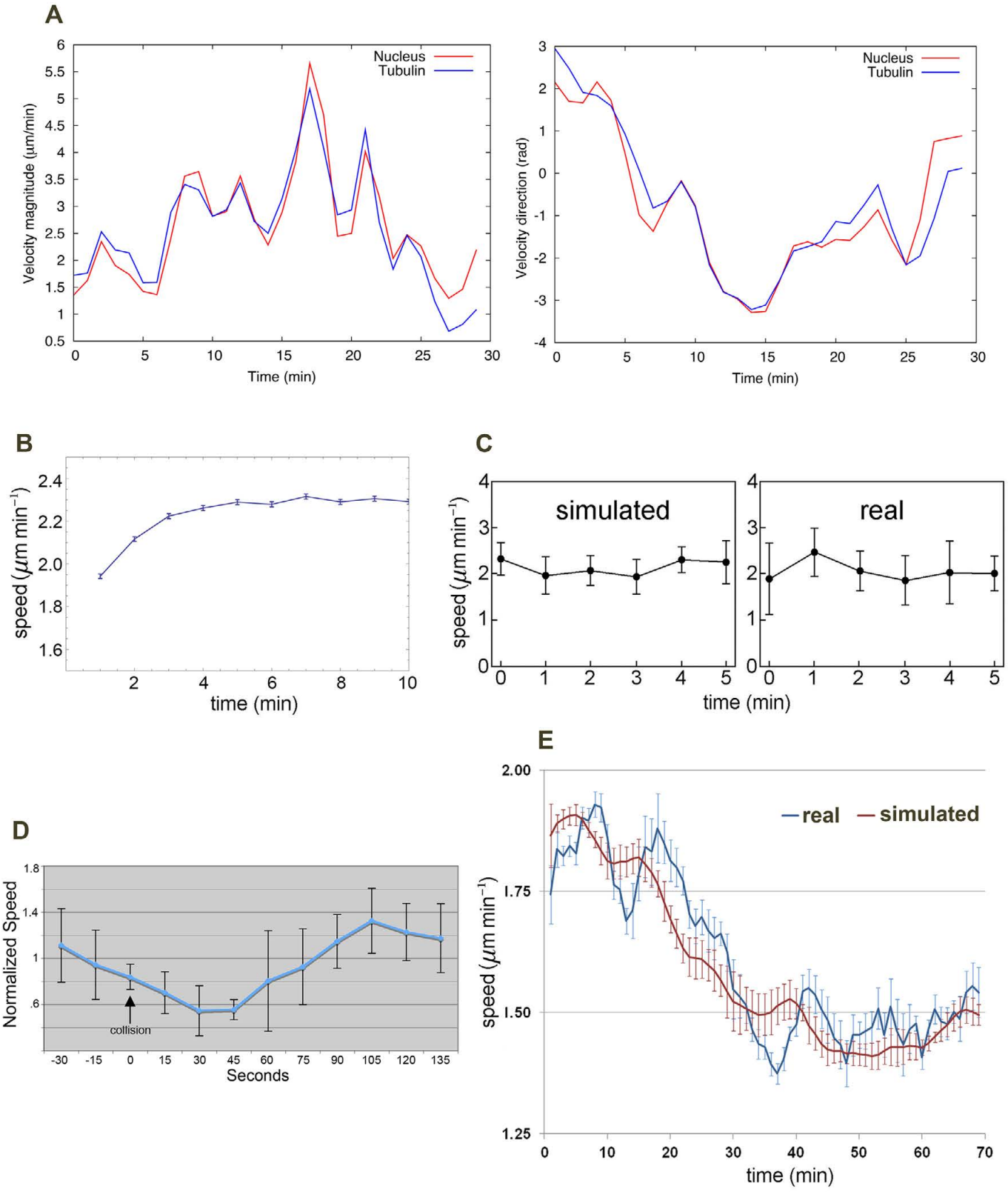


**Fig. S1. Kinematic testing of cell tracking and model algorithms.** (**A**) A cell was tracked for ~30 minutes using cell centroid or nuclear tracking methods to reveal that there was little difference between the magnitude or direction of velocity vectors using either method. (**B**) Approach to stationarity of the free motility model. Mean and standard error of the speed during the first 10 minutes of 10,000 simulations starting from rest. (**C**) Six successive mean (± s.e.m.) cell speeds for samples of ten non-colliding cells from the model simulations (left) and real haemocytes. The mean speeds are a good match but the real cells show slightly more variation. (**D**) Haemocyte collisions were observed over 15-second intervals for a total of 13 time points (four timepoints before the collision and nine after). Each individual track was smoothed over a 30-second sliding window and normalised for its average speed. As can be seen from the graph, haemocytes significantly decrease their speeds immediately following a collision and rapidly return to normal. Error bars indicate s.d. ($n$=6). (**E**) Mean haemocyte speed over time in real and simulated embryos. Cells initially have a high speed as they leave the midline and subsequently decrease their speed as they become constrained by neighbouring cells. Graphs are an average of three real and simulated embryos. Error bars represent s.e.m.

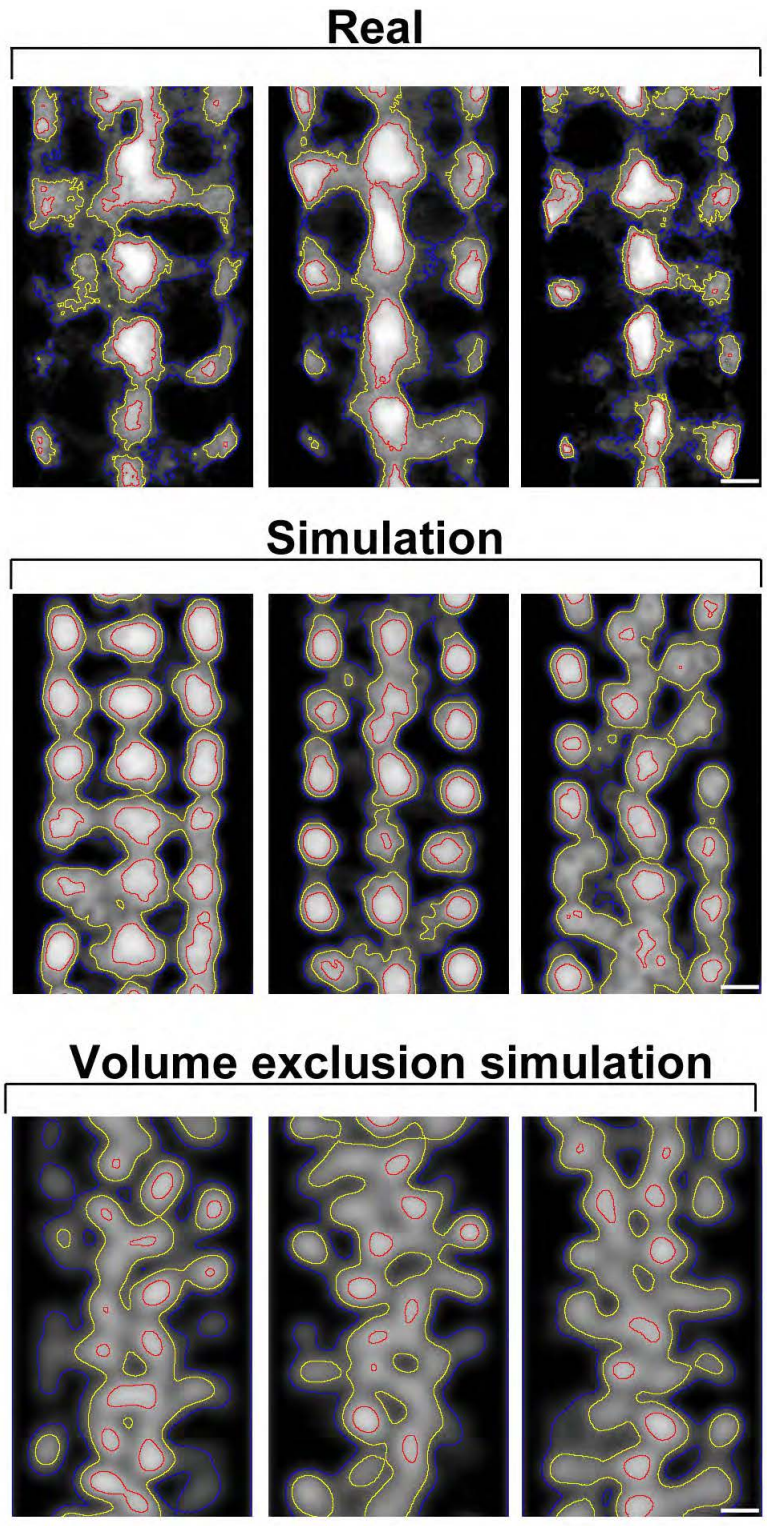


**Fig. S2. Example contour plots of domain maps from real and simulated embryos.** Note the subtle variation in final pattern in real and simulated embryos compared with a complete failure of pattern formation in the volume exclusion model. Scale bars: 15 μm.

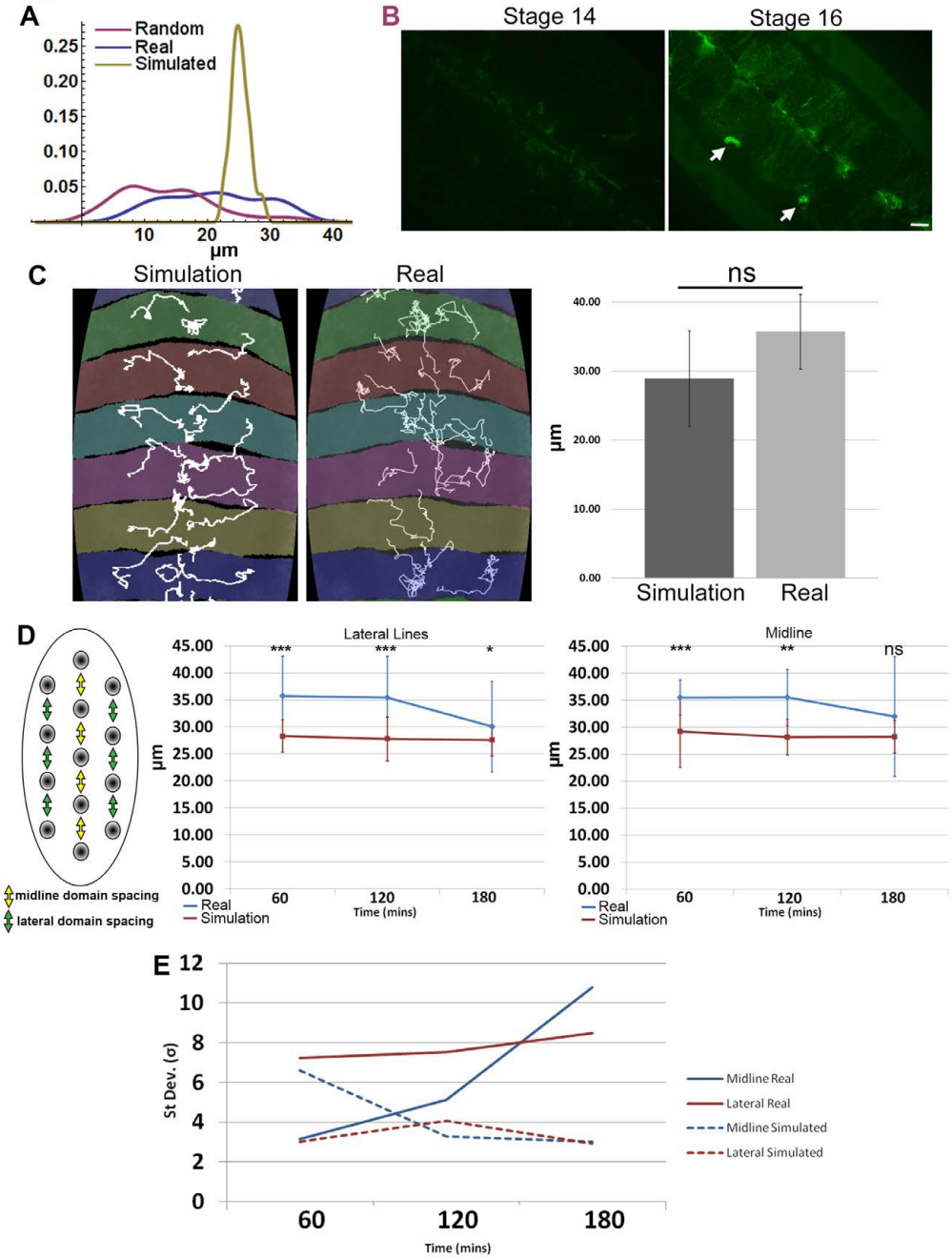


**Fig. S3. Examination of crosstalk with extrinsic cues.** (**A**) Probability distributions of nearest neighbour distances of real and simulated haemocytes, versus an equal number of randomly distributed cells. The median nearest neighbour distance in reality was slightly smaller than in simulations (20.43±7.85 versus 25.08±1.45 µm, respectively; $P<0.0001$, df=168, Mann-Whitney test), but significantly greater than a random distribution (20.43±7.85 versus 12.99±7.15 µm, respectively; $P<0.0001$, df=158, Mann-Whitney test). (**B**) Viking-GFP (Collagen IV) enhancer trap expression in stage 14 and 16 embryos. Arrows highlight haemocytes expressing high levels of the enhancer trap. Scale bar: 10 µm. (**C**) (Left) Haemocyte tracks in simulated and real embryos were overlaid onto a colour-coded map defining embryonic segments. (Right) Quantification of the percentage of tracks per embryo that crossed segment boundaries revealed that there is no significant difference between real and simulated embryos (35.70±4.40% versus 28.89±5.66%, respectively; $P>0.25$, df=4, unpaired *t*-test). (**D**) (Left) Schematic embryo domain map showing how midline and lateral line domain distances were calculated. (Right) Distances between domains residing within the midline and lateral lines were measured and plotted over time in real and simulated embryos. In reality, during the early stages of haemocyte dispersal, the mean distance between the domains was slightly greater than in the model (for the midline domains at 60 minutes 35.48±3.25 µm versus 29.17±6.62 µm, df=47, and 120 minutes 35.50±5.24 µm versus 28.15±3.28 µm, df=44. For the lateral domains at 60 minutes 35.72±7.38 µm versus 28.29±3.02 µm, df=87, and 120 minutes 35.44±7.67 µm versus 27.81±4.06 µm, df=93). However, at later stages the mean distances become less significant (for the midline at 180 minutes 31.95±11.06 µm versus 28.20±3.02 µm, df=49, and the lateral lines at 180 minutes 30.07±8.40 µm versus 27.55±2.91 µm, df=94). All tests used the Welch's correction *t*-test. (**E**) The standard deviations of the mean distances between domains within the midline and within lateral lines in real and simulated embryos were plotted over time. In real embryos, the standard deviation in midline domains was significantly less than domains in lateral lines early in haemocyte dispersal [60 minutes; two-sided Fisher's variance ratio test: variance ratio=0.16; df=(21,35), $P<0.0001$]. By contrast, in simulated embryos in early haemocyte dispersal (60 minutes) the standard deviation in midline domains was significantly greater than domains in lateral lines [two-sided Fisher's variance ratio test: variance ratio=4.67; df=(26,52), $P<0.0001$]. However, at later stages of haemocyte dispersal (180 minutes), the standard deviation in reality and simulated embryos was not significantly different [two-sided Fisher's variance ratio test: variance ratio=1.37; df=(26,43), $P>0.05$ and variance ratio=1.10; df=(23,51), $P>0.05$, respectively].

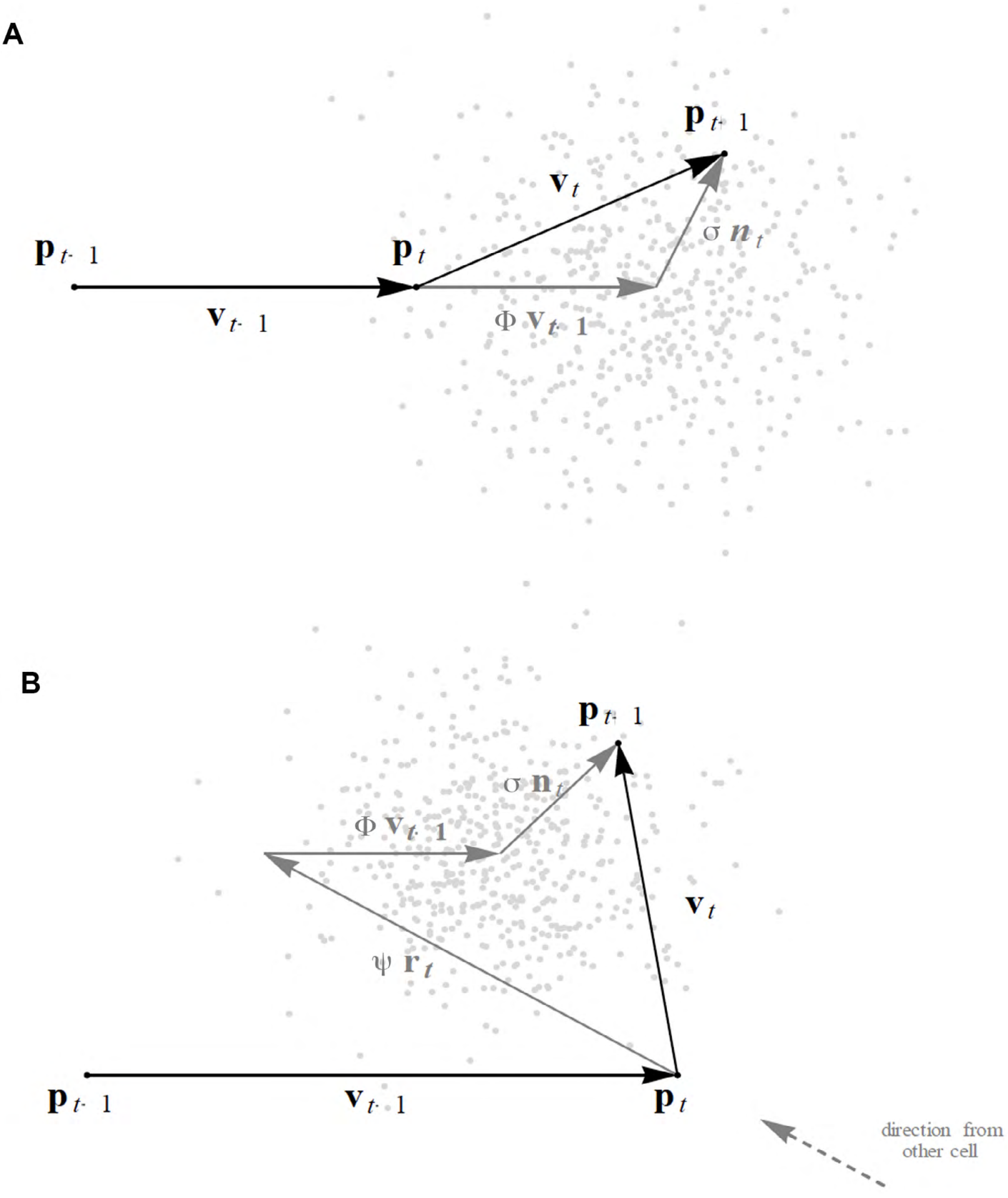


**Fig. S4. Graphical representation of the free-running and collision models.** (**A**) Free-running model. (**B**) Collision model. See Appendix S1 for details of the simulation.

Appendix S1

Simulation

The model consists of two alternatives and one of these is applied to each cell in turn at each time step of 1 min. The free-moving model is used when there are no other cells within collision distance of the cell being modeled and the collision model is used when the nearest neighboring cell is within the collision distance. In real cells the collision distance is remarkably constant at 30 μm. In the simulation, the motile cells start off by being confined to a narrow region located at the midline of the embryo. After a predetermined time interval, they are released to wander anywhere within the ventral embryo space. The boundary of this space is modeled as a chain of closely packed cells. These are stationary and do not take part in the simulation but they can act as neighboring cells during the collision response. This arbitrary boundary condition is partially justified by the observation that cells in the real embryo appear to be largely confined to the ventral space by contact interactions with other cells at the edge of the space.

The simplest 2-D discrete random process that we considered as a model for cell tracks, sampled at unit time intervals, is based on the recursive velocity process… $v_t = \phi v_{t-1} + \sigma n_t$ (Eq.1)… where the v are vector velocities, $\phi$ is a scalar persistence factor, $\sigma$ is a scalar standard deviation multiplier and n is a 2-D random noise vector with Gaussian distribution, mean = 0 and standard deviation = 1 on both axes. For $0 <= \phi < 1$ the process becomes stationary and can be used to generate the successive position vectors, p, in a 2-D random walk: $p_{t+1} = p_t + v_t$. The model is represented graphically in supplementary material Fig. S4A, where the cloud of grey dots is a sample of the possible positions of $p_{t+1}$.

To estimate $\phi$, we extracted all pairs of successive velocities ($v_{t-1}$, $v_t$) and rotated each pair so that $v_{t-1}$ points along the +x axis. We then take the x-components of the rotated vectors so we

can reduce Eq.1 to the scalar equation $vx_t = \phi vx_{t-1} + \sigma nx_t$. Taking the expectations of these quantities we get $E[vx_t] = \phi E[v_{t-1}] + \sigma E[nx_t]$. Since we know $E[nx_t] = 0$, the estimate of $\phi$ becomes $\phi_{estimate} = Mean[vx_t]/Mean[vx_{t-1}]$. To estimate $\sigma$, we plot the noise $\sigma n_t = v_t - \phi v_{t-1}$ using the rotated vectors and the estimated $\phi$ and take standard deviations on both axes. We thus get two estimates which should not be significantly different if the model is valid: $\sigma x_{estimate} = SD[vx_t - \phi vx_{t-1}]$. $\sigma y_{estimate} = SD[vy_t - \phi vy_{t-1}]$.

When a collision occurs at time t we need a new term to take account of the direction of the other cell (since we will show that this is significant): $v_t = \psi r_t + \phi v_{t-1} + \sigma n_t$ … (Eq.2)…. Here $r_t$ is a unit vector pointing to $p_t$ from the position of the other cell at the time of collision, $\psi$ is a new scalar factor that determines the influence of the other cell's position. $v_t$ is now the vector velocity during the first time interval after collision and $\phi$ and $\sigma$ may take on new expected values. The model represented graphically in supplementary material Fig. S4B, where the cloud of grey dots is a sample of the possible position of $p_{t+1.}$ To estimate the parameters, we extracted all pairs of successive velocities ($v_{t-1,}$ $v_t$) from colliding cells that had first made contact at time t and again rotated each pair so that $v_{t-1}$ points along the +x axis. We then take the y-components of the rotated vectors so we can reduce Eq.2 to the scalar equation $vy_t = \psi ry_t + \phi vy_{t-1} + \sigma ny_t$. Taking the expectations of these quantities we get $E[vy_t] = \psi E[ry_t] + \phi E[vy_{t-1}] + \sigma E[ny_t]$ which becomes $E[vy_t] = \psi E[ry_t]$ since two terms vanish. Now the expectations of $ry_t$ and hence of $vy_t$ are zero if the target cells are distributed equally to the right and left of the approaching cells. However, assuming the cell behavior is left/right symmetrical, we can bias the distribution by taking mirror images of all the tracks where the target cells is, say, on the left (so $ry_t > 0$ in all cases). The estimate of $\psi$ then becomes $\psi_{estimate} = Mean[vy_t]/Mean[ry_t]$. Considering the x-components we have $E[vx_t] = \psi E[rx_t] + \phi E[vx_{t-1}] + \sigma E[nx_t]$ which becomes $E[vx_t] = \psi E[rx_t] + \phi E[vx_{t-1}]$. Using the estimate of $\psi$, the estimate of $\phi$

then becomes: $\phi_{estimate}$ = Mean[$vx_t$ - $\psi rx_t$]/Mean[$vx_{t-1}$]. Finally, to estimate $\sigma$, we plot the noise $\sigma n_t = v_t - \psi r_{t} - \phi v_{t-1}$ using the rotated vectors and estimated $\phi$ and $\psi$ and take standard deviations on both axes.

Thirty cells were chosen as the typical number of hemocytes on the ventral surface, and the cells were simulated in rotation at each 1-minute interval using *Mathematica*©. Each cell was simulated using the free motility model unless its nearest neighbor was within 30 μm, in which case the collision model was used. The boundary of the simulated embryo consisted of 70 equally spaced points, which would also invoke the collision model.