Supplementary Materials and Methods

The reaggregation simulation model consists of a set of ureteric and non-ureteric cells represented as points in a three dimension rectangular volume, with adhesions between certain pairs of ureteric cells. Cells are initially placed at random. At each subsequent time point, positions are updated based on repulsive impulses between pairs of nearby cells and attractive impulses between pair of ureteric cells connected by an adhesion, plus a random impulse at each cell. At each time step, prior to the calculation of these impulses, the adhesion relationships are updated by removing any adhesions between cells separated by more than a threshold distance, and adding, with a certain probability, an adhesion between each pair of ureteric cells which are separated by less than a second (smaller) threshold distance but do not currently have an adhesion relationship. After cell positions have been updated based on the calculated impulses, positions beyond the defined volume are reverted to the nearest boundary point, in order to maintain a known cell density.

Note that physical momentum is not included in the model. It is assumed that the external forces acting on a cell rapidly reach equilibrium with drag, and hence cell velocity at any time can be well approximated as a function of the instantaneous forces acting on the cell.

Model specification

The simulation parameters are specified in the following table, with corresponding variable names used in the code implementing the algorithm.

parameter	variable name	description	Value
σ_{sim}	brownianMotionPerTimestepStdev	The standard deviation of the random displacement added to each dimension of the position of each cell at each time step.	0.2
n	type1cells + type2cells	Total number of cells	6800 / 13600
m	type2cells	Number of ureteric cells	6800
X, Y, Z	xDim,yDim,zDim	Define 3D volume in which the simulation takes place	680,250,200
d_r	cellSpacingForceDist	Repulsion force applied to cells closer than this distance, modelling physical exclusion	8
f_r	cellSpacingForceStrength	Repulsion force is proportional to this value	0.8
d_a^c	createAdhesionDist	Adhesion bonds may be formed between cells within this distance of each other.	10
p_a	adhesionFormationProbability	If they are close enough to form an adhesion bond, this is the probability of a bond being created each time step.	0.005
d_a^b	breakAdhesionDist	Adhesion bonds between cells further apart than this are removed	12

d_a	adhesionSpringLength	If there is an adhesion bond between two cells but they are greater than this distance, an attractive force is applied	8
f_a	adhesionSpringForce	Adhesion force is proportional to this value	0.02

The model representation at time point $t \in \mathbb{Z}^+ \cup \{0\}$ consists of a set of n cell positions $\boldsymbol{p_i}(t) \in \boldsymbol{V} = [0, X] \times [0, Y] \times [0, Z]$, for $i = 1, \ldots, n$, of which the first $m \leq n$ correspond to ureteric cells capable of adhering to each other, and a set $A(t) \subseteq \{(i, j): 1 \leq i < j \leq m\}$ representing adhesions between ureteric cell pairs.

Let $\mathcal{N}_i(t) = \{j: (i, j) \in A(t) \; or \; (j, i) \in A(t)\}$ represent the set of cells adhering to cell i at time t, and let $d_{ij}(t) = \|\boldsymbol{p}_i(t) - \boldsymbol{p}_j(t)\|$ and $\hat{\boldsymbol{p}}_{ij}(t) = \left(\boldsymbol{p}_i(t) - \boldsymbol{p}_j(t)\right)/d_{ij}(t)$. The model is initialised by generating n cell positions selected uniformly at random from the volume $\boldsymbol{V}$: $\boldsymbol{p_i}(0) \sim \mathcal{U}(0, X) \times \mathcal{U}(0, Y) \times \mathcal{U}(0, Z)$.

At time step $t \geq 1$, the model is updated as follows:

$A(t) = \{(i, j) \mid (i, j) \in A(t-1) \wedge d_{ij}(t) \leq d_a^b\} \cup \{(i, j) \mid 1 \leq i < j \leq m \wedge d_{ij}(t) \leq d_a^c \wedge \mathrm{I_{ijt}}\}$,

where I_{ijt} is a independent random boolean variable with $P(I_{ijt}) = p_a$;

$\boldsymbol{p}_i(t)' = \boldsymbol{p}_i(t-1) + \boldsymbol{F}_b(i, t) + \sum_{j \neq i} \boldsymbol{F}_r(i, j, t) + \sum_{j \in \mathcal{N}_i(t)} \boldsymbol{F}_a(i, j, t)$, where

$\boldsymbol{F}_r(i, j) = \max(d_{ij}(t-1) - d_r, 0) \cdot f_r \cdot \hat{\boldsymbol{p}}_{ij}(t-1)$,

$\boldsymbol{F}_a(i, j) = \min(d_{ij}(t-1) - d_a, 0) \cdot f_a \cdot \hat{\boldsymbol{p}}_{ij}(t-1)$,

$\boldsymbol{F}_b(i, t) \sim \mathcal{N}(0, {\sigma_{sim}}^2)^3$;

$$\boldsymbol{p}_i(t) = \begin{pmatrix} \max(0, \min(\boldsymbol{p}_{i1}(t)', X)) \\ \max(0, \min(\boldsymbol{p}_{i2}(t)', Y)) \\ \max(0, \min(\boldsymbol{p}_{i3}(t)', Z)) \end{pmatrix}$$

Simulation calibration

Simulation parameters were selected to match the experimental data in cell density (cells per unit volume), speed and separation between neighbouring cells.

Brownian motion (BM) is isotropic, with each component of motion independent, and thus the calibration calculations reduce to one dimension. Let $x(t)$ be the position of a cell at time point t along some arbitrary fixed axis, and $v(t) = x(t) - x(t-1)$. The velocity of a cell under BM is uncorrelated between successive time points, and can be characterised by a

diffusion constant $D = E((x(t+t_0) - x(t_0))^2)/t$. Live imaging analysis indicated significant positive autocorrelation between successive 15min time steps for Six2 cells, but motion over longer time scales that is consistent with BM for both Six2 and Hoxb7 cells (see Fig. 2). The observed single time-step displacement is modelled as $v(t) = f(t) + \lambda v(t-1)$, where $f(t) \sim \mathcal{N}(0, \sigma_f^2)$ is an independent random variable and λ is the autocorrelation between successive time steps. Then

$$(x(t+t_0) - x(t_0))^2 = \left[\sum_{i=1}^{t} v(i+v_0)\right]^2 = \sum_{i=1}^{t}\sum_{j=1}^{t} v_x(i+t_0)v_x(j+t_0)$$

and hence

$$E((x(t+t_0) - x(t_0))^2)/t = \sum_{i=1}^{t}\sum_{j=1}^{t} \lambda^{|i-j|}\,\sigma^2/t = \sigma^2(t + (t-1)\lambda + (t-2)\lambda^2 + \cdots)/t,$$

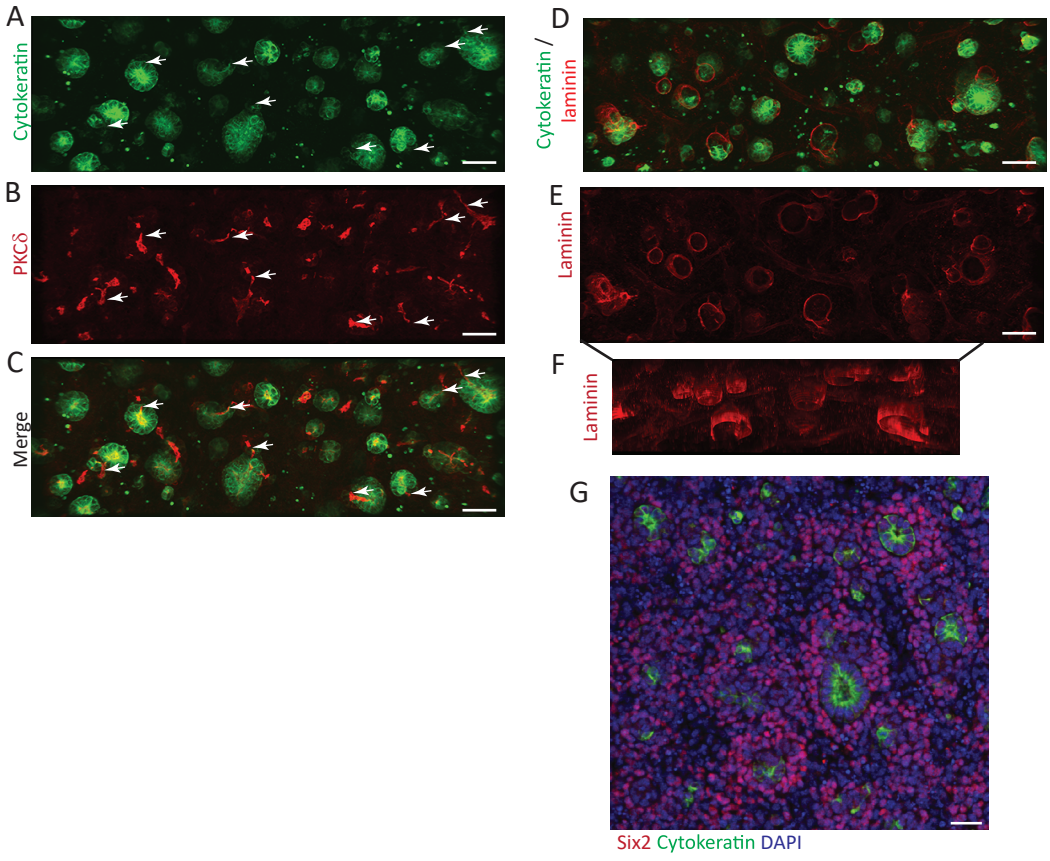
where σ^2 is the variance of $v(t)$. Since $0 < \lambda < 1$, for large t this approximates $\sigma^2/(1-\lambda)$. We estimate $\lambda = R(1)$ and σ^2 from the live imaging data. A complication is that the observed single time step cell displacements $v_x(c,t)$ and $v_y(c,t)$, while broadly following a normal distribution, contained a number of large outliers which appear to be due to errors in the Imaris spot tracking algorithm. Absolute values above 5µm were removed. The variance was calculated across all remaining values of $v_x(c,t)$ and $v_y(c,t)$, giving 2.43 and 2.01 µm^2/time step respectively for Hoxb7 and Six2 cells. Single time step autocorrelation on the Six2 data with outliers removed was 0.11. Applying the autocorrelation correction yielded diffusion rates of 18.2 and 16.9 µm^2/hour (7.5 time steps per hour). The simulation parameter σ_{sim} was set to 0.2, in order to provide good computational performance combined with numerical stability, implying a diffusion rate of 0.04 µm^2/simulation time step. Simulated and observed diffusion can thus be matched by setting 455.8 (Hoxb7) or 421.5 (Six2) simulation time steps per hour. The focus of the simulation is ureteric cell behaviour, so a conversion of 450 simulation time steps per hour was used.

The simulation volume was 3.4×10^7 µm^3, with $(X,Y,Z) = (680,250,200)$. The x and y dimensions were chosen to match the experimental image section dimensions, while the z dimension was chosen to be large enough to avoid edge effects and provide a substantially greater volume than the aggregate experimental data, to ensure that random variability in the simulation results was negligible compared to experimental data. We simulated 6800 ureteric

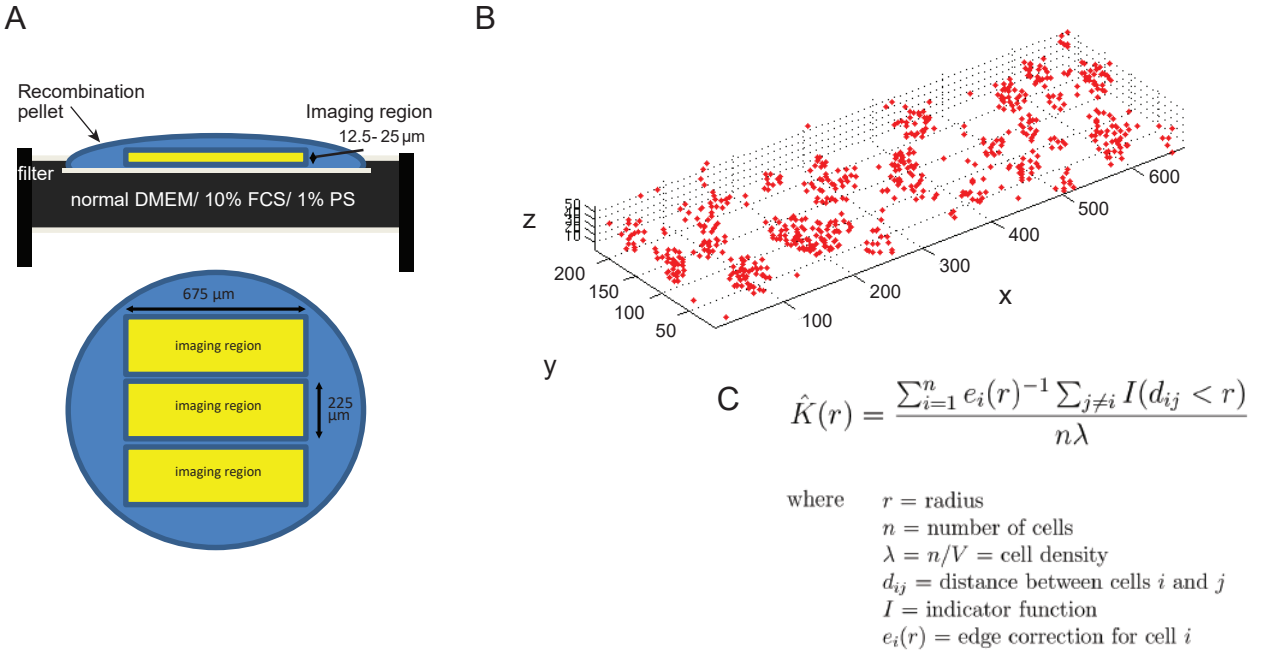
cells in this volume, corresponding to a density of 2×10^{-4} cells/μm^3, calibrated to the approximate average density of identified Cytokeratin$^+$ cells in the timecourse data (Fig. 3E).

Cell interaction effects were calibrated to allow the formation of stable cell aggregations while ensuring that the cell repulsion force is sufficient to maintain proper cell spacing. Poor calibration of spacing distance will result in incorrect cluster density, while poor balance between adhesive and repulsive forces may result in an unrealistic collapse of larger simulated clusters. Either effect will bias the simulated progression of clustering, as the probability of collision between clusters is related to volume. The calibration of cell interaction forces was informed by the measured distances between Cytokeratin$^+$ cells (Fig. S2D), but based primarily on plots of cluster size (number of cells) versus cluster volume (Fig. S2E). This is to ensure realistic cluster density across the full range of cluster sizes. Nearest neighbour distances are more variable for experimental than simulated data, possibly due to biological variability or imprecision in image analysis, but this has little effect.

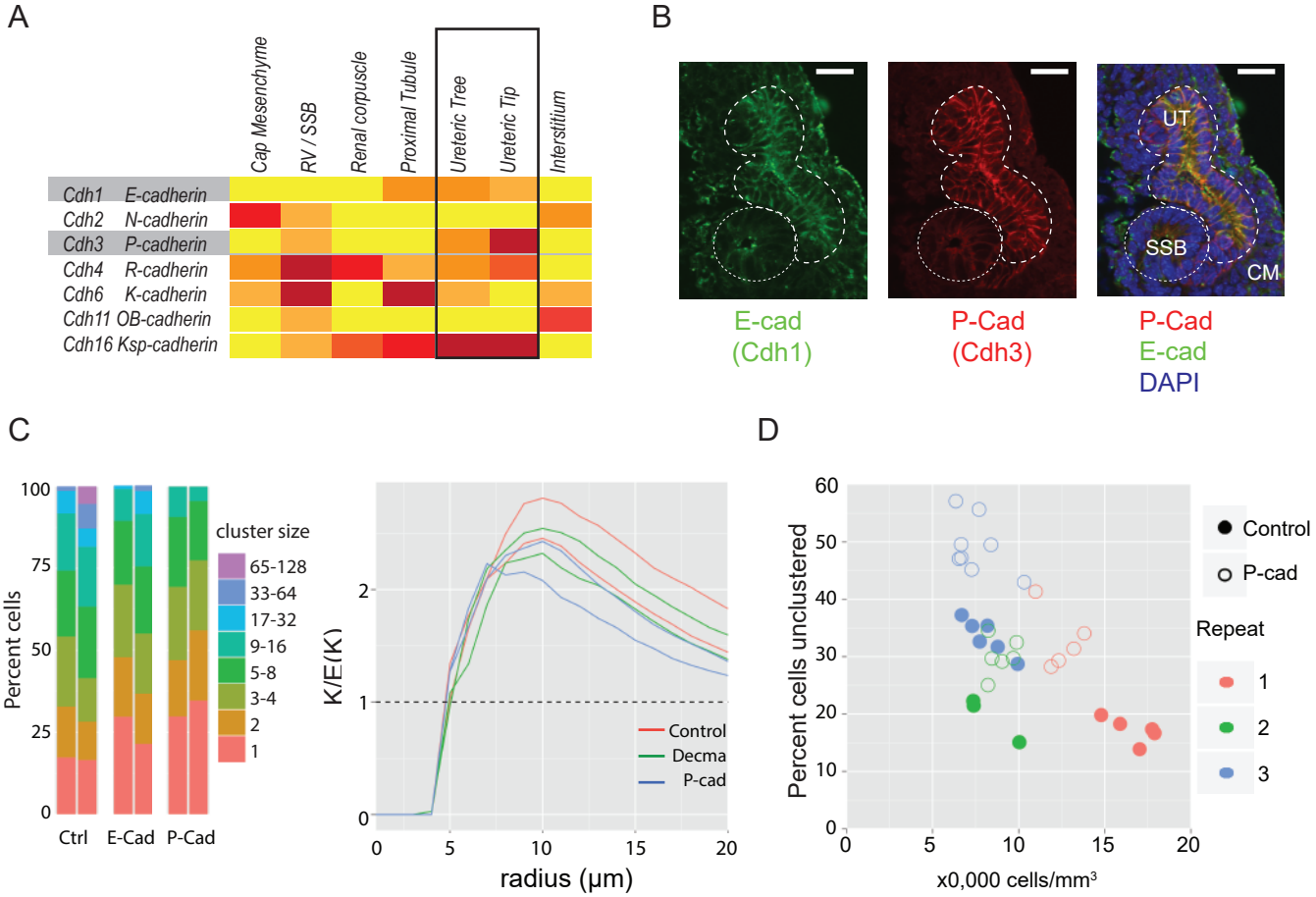
Supplementary Figures



Supplementary Figure 1. Low magnification images of morphogenetic events within reaggregation cultures. A-C. Immunofluorescent labelling of early nephron formation after 24 hours of culture. **A**. UE, cytokeratin – green; **B.** Apical surface of epithelium of UE and nephron, PKCδ –red. **C.** Merge. White arrows indicate evidence of nephrons with connections to UE based upon apical staining of PKCδ. Scale bar 100μm **D-F.** Formation of basement membrane around UE and attached early nephron. **D.** UE, cytokeratin (green) and pan laminin-labelled basement membrane (red). **E.** Pan laminin (red). **F.** Z-stack of pan-laminin staining illustrating intact basement membrane around UE and forming nephrons. Scale bar 100μm. **G.** Relative localisation of UE and CM cells 24 hours post reaggregation. UE (cytokeratin; green); CM (Six2; red). Scale bar 50μm

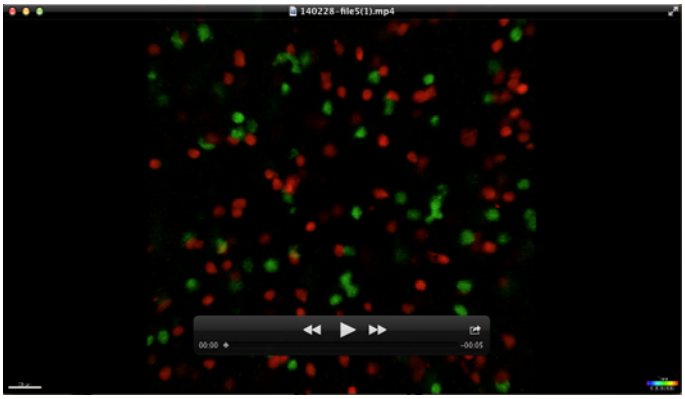

A
Recombination pellet
Imaging region 12.5- 25 µm
filter
normal DMEM/ 10% FCS/ 1% PS
675 µm
imaging region
imaging region
imaging region
225 µm
B
z
x
y
C
$\hat{K}(r) = \frac{\sum_{i=1}^{n} e_i(r)^{-1} \sum_{j \neq i} I(d_{ij} < r)}{n\lambda}$
where
r = radius
n = number of cells
$\lambda = n/V$ = cell density
d_{ij} = distance between cells i and j
I = indicator function
$e_i(r)$ = edge correction for cell i
D
Cytokeratin+ cell spacing
Proportion of cells
Nearest neighbour distance (µm)
hours culture
2
8
12
24
48
E
Cluster density
Experimental timecourse
Simulation
Cluster volume (µm³)
Cells per cluster
Cells per cluster
F
Sensitivity analysis
Standard settings
No non-adhesive cells
25% fewer UE cells
Immediate adhesion
1/10th adhesion
Mean cluster size
Hours reaggregation

Supplementary Figure 2. Quantification of clustering within reaggregation cultures and model calibration and sensitivity analysis. A. Diagram of how an explant was sampled for quantification. **B.** Sample RipleyGUI output. Cytokeratin^{+} cell positions identified in an image region from a 24hr reaggregation and imported into RipleyGUI for $K(r)$ calculation. **C.** Ripley's K formula, estimating the average number of other cells within distance r of a reference cell, divided by overall cell density (Hansson et al, 2013). **D.** Distribution of nearest neighbour distances in Cytokeratin timecourse data, used to select the clustering threshold of $12\mu m$. As clustering progresses, the proportion of cells at distances above this threshold falls consistently. This threshold was used to identify clusters in the experimental data and was also used in the calibration of simulated cell interaction forces. **E.** Scatter-plot of cluster size (measured in number of cells) against volume, experiment and simulation. The relationship is similar for the experimental and simulation data. This was used for the calibration of simulated cell interactions (adhesion, and repulsion due to spatial exclusion). **F.** Consistency and sensitivity analysis. Plots of mean cluster size versus simulated reaggregation time, for standard settings and 4 variants, each with three runs to demonstrate consistency. Clustering is insensitive to the presence of non-adhesive cells, but sensitive to the density of adhesive cells. This may explain most experimental variability in clustering. The propensity of UE cells to form adhesive bonds could not be directly measured, and clustering is also sensitive to large variation in this parameter.

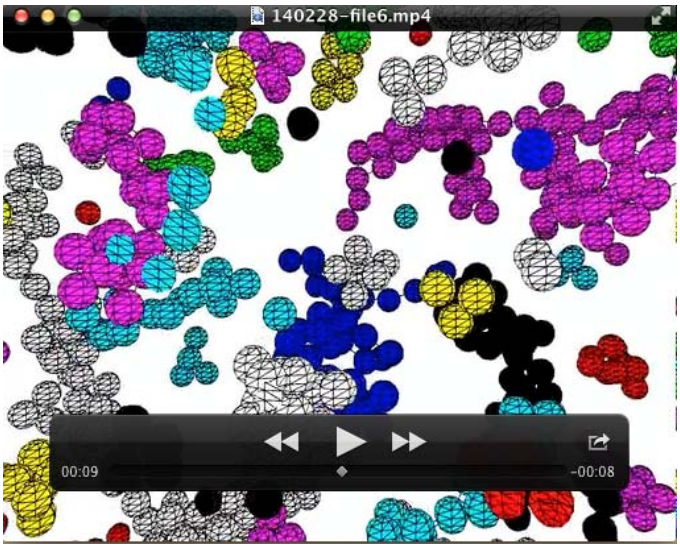


Supplementary Figure 3. Comparative distribution of cadherins during kidney development and effect of blocking antibodies to E-cadherin and P-cadherin on self-organisation. A. Expression of the major cadherins present in the mouse developing kidney displayed as a heat map of gene expression across 7 cell compartments. Cadherins are listed by gene name and common protein name. Box highlights expression within the ureteric epithelial compartments. Data was extracted from www.gudmap.org. RV, renal vesicle; SSB, S-shaped body. **B**. Immunofluroescence of E-cadherin (DECMA antibody, green) and P-cadherin (red) in cortical sections of the 13.5dpc developing kidney. CM, cap mesenchyme; SSB, S-shaped body; UT, ureteric tip. DAPI (blue). Scale bar 30μm. **C.** DECMA and P-cad results at 8hr as K/E(K) and cluster sizes. Strongest reduction in clustering produced by p-cad inhibition. **D.** Plots of cell density versus proportion of unclustered UE cells between control (solid circles) and anti-P-cadherin antibody treated (open circles) cultures from 3 different 8hr experiments. In all instances, anti-P-cadherin antibody treated cultures show a higher number of unclustered UE cells. **E, F**. Sample images from P-Cad inhibition experiment showing control (E) and experimental (F) groups at 8 and 24 hours. Immunofluorescence for cytokeratin (green) and Six2 (red) was used to identify UE and CM-derived cells respectively; DAPI (blue) marks nuclei. Scale bar 30μm.

Supplementary movies:



Movie 1. Timelapse imaging of reaggregation culture showing ureteric and cap mesenchyme cells. The movie depicts a single optical slice imaged across a 7 hour period commencing 2 hours after reaggreation. Red cells represent Six2-tdTomato cap mesenchymal cells. Green cells represent Hoxb7-GFP+ ureteric epithelial cells. Note that the green cells move out of field with time.



Movie 2. Clustering simulation based on a model of differential adhesion. Shows 48 hours of simulated clustering, sampled every 20 time steps. Only ureteric epithelium (adhesive) cells in a portion of simulation shown. Random colours are used to distinguish clusters, and similarity in colours between clusters has no meaning. See Fig. 4A, Fig. S2D-F and online methods for description and derivation of model. A cluster is defined as a set of cells where any two cells are connected by a path consisting of one or more adhesion bonds. Colour changes mark the joining of clusters; for example at 15 seconds the main black and blue clusters merge.