

<https://doi.org/10.1038/s41540-025-00502-4>

# An integrative phenotype-structured partial differential equation model for the population dynamics of epithelial-mesenchymal transition



Jules Guilbeteau<sup>1,4</sup>, Paras Jain<sup>2,4</sup>, Mohit Kumar Jolly<sup>2</sup>✉, Camille Pouchol<sup>3</sup>✉ & Nastassia Pouradier Duteil<sup>1</sup>✉

Phenotypic heterogeneity along the epithelial-mesenchymal (E-M) axis contributes to cancer metastasis and drug resistance. Recent experimental efforts have collated detailed time-course data on the emergence and dynamics of E-M heterogeneity in a cell population. However, it remains unclear how different intra- and inter-cellular processes shape the dynamics of E-M heterogeneity. Here, using Cell Population Balance model, we capture the dynamics of cell density along E-M phenotypic axis resulting from interplay between—(a) intracellular regulatory interaction among biomolecules, (b) cell division and death and (c) stochastic cell-state transition. We find that while the existence of E-M heterogeneity depends on intracellular regulation, heterogeneity gets enhanced with stochastic cell-state transitions and diminished by growth rate differences. Further, resource competition among E-M cells can lead to both bi-phasic growth of the total population and/or bi-stability in the phenotypic composition. Overall, our model highlights complex interplay between cellular processes shaping dynamic patterns of E-M heterogeneity.

Intra-tumour heterogeneity—the co-existence of multiple distinct cellular phenotypes in a tumour—is being increasingly reported to associate with poor patient outcomes. It contributes to both metastasis and therapy resistance—two major unsolved clinical challenges<sup>1,2</sup>. Such heterogeneity can arise at a genetic level over the course of tumour evolution and manifests as different clonal populations<sup>3–5</sup>. However, over the last two decades, non-genetic or phenotypic heterogeneity among cancer cells has been identified as a key driver of disease aggressiveness<sup>6–9</sup>. Such heterogeneity is often characterised by single-cell measurements (flow cytometry, mass cytometry, RNA-seq, ChIP-seq, ATAC-seq), showing diversity among cells in a population at the proteome, transcriptome and epigenome levels. A canonical example of non-genetic heterogeneity is along the Epithelial-Mesenchymal (E-M) phenotypic states in carcinomas<sup>10,11</sup>. E-M heterogeneity has been implicated in poor patient prognosis, and several in vitro and in vivo studies have shown its contribution to cancer metastasis and drug resistance<sup>12–15</sup>. Thus, understanding cellular mechanisms driving E-M heterogeneity can help improving therapeutic outcomes.

An iterative interplay among in silico and in vitro studies has contributed enormously to deciphering how non-genetic heterogeneity emerges in a population. Mathematical modelling of regulatory networks underlying Epithelial-Mesenchymal Plasticity (EMP) have reported the co-existence of multiple cellular phenotypes—Epithelial (E), Mesenchymal (M) and one or more hybrid E/M (H) cell states<sup>16–20</sup>. Their co-existence has been experimentally reported in varying ratios in multiple cell lines and primary tumours<sup>10,21</sup>. Synthetic perturbation of the underlying regulatory network led to the loss of bimodal nature of canonical epithelial markers such as E-cadherin and altered the phenotypic distribution<sup>12,13,22,23</sup>. The relative stability of cells in different phenotypes, and consequently the phenotypic distribution of the population at a given time is governed by alternative, self-reinforcing network interactions between epithelial and mesenchymal players in the gene regulatory network<sup>24</sup>. The epigenetic regulation of genes further reinforces transcriptional state. Happening at timescales slower than transcriptional regulation, the epigenetic regulation can be reversible or irreversible, as observed experimentally. Thus, epigenetic remodelling can lock cells transiently or permanently in a cell-state, impacting the population

<sup>1</sup>Sorbonne Université, CNRS, Université Paris Cité, Inria, Laboratoire Jacques-Louis Lions (LJLL), Paris, France. <sup>2</sup>Department of Bioengineering, Indian Institute of Science, Bangalore, India. <sup>3</sup>Université Paris Cité, FP2M, CNRS FR 2036, MAP5 UMR 8145, Paris, France. <sup>4</sup>These authors contributed equally: Jules Guilbeteau, Paras Jain. ✉e-mail: [mkjolly@iisc.ac.in](mailto:mkjolly@iisc.ac.in); [camille.pouchol@u-paris.fr](mailto:camille.pouchol@u-paris.fr); [nastassia.pouradier\\_duteil@sorbonne-universite.fr](mailto:nastassia.pouradier_duteil@sorbonne-universite.fr)

phenotypic distribution<sup>25,26</sup>. Further, cell-to-cell communications either through neighbourhood interactions or paracrine signalling also reshapes the phenotypic distribution<sup>27–30</sup>. Hence, a diversity of regulatory interactions within and among cells can contribute to shaping the E-M heterogeneity patterns in a cell population.

Another set of factors such as asymmetric cell division<sup>31,32</sup>, stochastic biochemical noise<sup>33</sup>, differences in cellular microenvironment<sup>34</sup> and variable cell-cycle dynamics<sup>35</sup> can amplify dynamic heterogeneity in a cellular population. These factors alter cell-to-cell variability in protein levels in a population, thus contributing to their functional heterogeneity (differential activation of signalling pathways) when these cells are exposed to cytotoxic or EMT-inducing growth factors<sup>36,37</sup>. Despite extensive efforts in investigating above-mentioned regulatory and stochastic processes in E-M plasticity and heterogeneity, only a few computational models have incorporated these processes within a growing and dividing heterogeneous cellular population.

In the present work, we use Cell Population Balance modelling to study the dynamics of E-M heterogeneity in a growing cell population, relying on partial differential equations (PDEs). In general, Cell Population Balance models capture the evolution of the cell density given the regulatory, stochastic, and population dynamics of cells without dealing with individual cell level information. Our Cell Population Balance model combines three main cellular processes: (1) *growth* due to cell division and cell death, (2) *state regulation*, which corresponds to the time-evolution of the regulatory molecules defining a cell's phenotype, based on corresponding ODE models and (3) *stochastic cell transition*, which aggregates all sources of stochasticity in the fate of a cell's phenotype.

With the developed model, we explore how heterogeneity along the E-M axis emerges at a population level. We first show that the Cell Population Balance model can capture the previously reported dynamical features of hysteresis, and epigenetic regulation during cells undergoing an Epithelial-Mesenchymal Transition (EMT) followed by a Mesenchymal-Epithelial Transition (MET). Next, with the full model incorporating state regulation, cell growth and transition, we report that increasing biochemical noise in regulatory molecules can enhance E-M heterogeneity given that the regulatory network operates in the multi-stable phenotypic regime. On the other hand, E-M heterogeneity can also be lowered down because of differences in subpopulations growth rates. Further, for an abrupt change in E-M phenotypic composition of a population, the timescale of re-equilibration depends on the rate at which intra-cellular constitutive EMT signal (here, SNAIL) re-equilibrates. And lastly, we show that resource competition for cellular growth among E-M subpopulations can give rise to biphasic growth or initial phenotypic composition dependent steady state E-M heterogeneity.

For all simulations, we use a scheme from the family of particle methods, based on solving a properly defined set of ordinary differential equations (ODEs)<sup>38</sup>. These are particularly adapted for models based on PDEs such as those developed in the present paper. For the derivation and analysis of particle methods in that context, we refer to the recent theoretical work<sup>39</sup>, where it is proved that solutions are properly approximated by the proposed numerical method.

## Results

### Development of a general Cell Population Balance model capturing cell state regulation, cell growth and cell state transition

In this work, we aim to describe a large population of cells which are characterised by their phenotypes. More precisely, for each cell, the quantity  $y \in \mathbb{R}^n$  represents the concentration of several different (internal or external) molecules associated with the cell. Here,  $n$  represents the dimension of the phenotypic space, that we will also refer to as 'state space'. The dimension  $n$  of the phenotypic space will then correspond to the number of different molecules that determine the phenotype of each cell. With this approach, the evolution of a population of  $N_c$  cells can be described by a system of  $N_c$  coupled Ordinary or Stochastic Differential Equations, giving the evolution of each cell's phenotype  $y(t)$ . As a consequence, the number of

equations to be solved grows with the number of cells in the population, and such a model can rapidly give rise to costly simulations.

Another approach, usually referred to as *macroscopic* or *continuous* approach, consists of focusing on the cell density instead of tracking the evolution of each individual cell's phenotype. The cell density that we will denote by  $u$  from here onwards, is then a function of the phenotype  $y$  and of time  $t$ . More specifically, for each subset  $\Omega \in \mathbb{R}^n$  of the phenotypic space, the quantity  $\int_{\Omega} u(t, y) dy$  represents the number of cells whose phenotype belong to  $\Omega$  at time  $t$ . The evolution of the cell density  $u$  is described by a Cell Population Balance model, which is, mathematically speaking, a PDE. This approach is known to be a good approximation of the microscopic model when the number of cells is large.

To give an intuition for the final PDE model, we consider a simplified case, in which a cell's phenotype is assumed to be determined by the concentration of one given molecule, so that the phenotypic space is one-dimensional ( $n = 1$ ). The function  $u$  is to be understood in the following way: for any phenotype interval  $[a, b]$ , the integral  $\int_a^b u(t, y) dy$  represents the number of cells whose phenotype belongs to  $[a, b]$ . Thus, the total number of cells is computed by  $\int_{\mathbb{R}} u(t, y) dy$ , which will be denoted by  $\rho(t)$ . In our model, the total number of cells  $\rho(t)$  is amenable to evolve in time.

Three mechanisms will be considered to participate in increasing or decreasing the number of cells within each interval  $[a, b]$ : growth, cell regulation and stochastic state transition. Their interplay is illustrated in Fig. 1 and individual descriptions are provided as follows:

The first mechanism that we include in our model is 'State Regulation'. At the level of a single cell, the evolution of its phenotype  $y(t)$  is assumed to be given by the following ODE:

$$\dot{y}(t) = f(t, y(t)), \quad (1)$$

where  $\dot{y}$  denotes the derivative of  $y$  with respect to time.

We extrapolate this simple ODE model to describe the evolution of the cell density of a whole population. In the PDE framework, the so-called *advection* term accounts for all cells whose phenotypes will enter or leave the phenotypic region  $[a, b]$  during a small time interval as a result of their inner evolution. Taking only this mechanism into account, the variation of the number of cells in the region  $[a, b]$  between two time instants  $t$  and  $t + \Delta t$  is computed as:

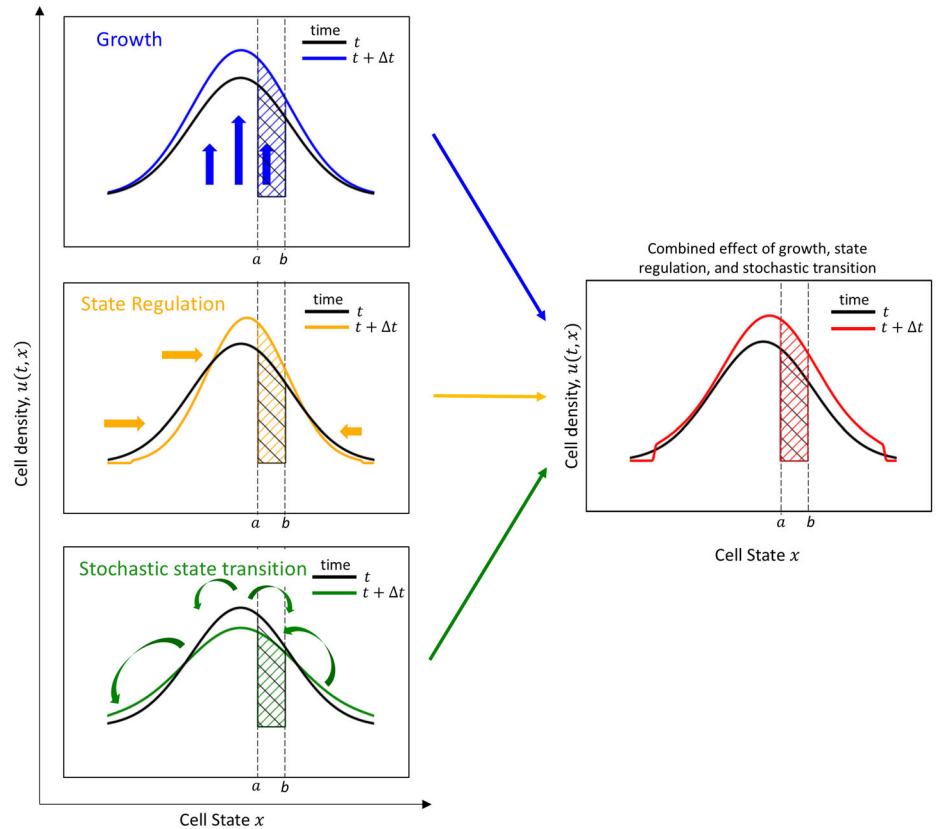
$$\int_a^b u(t + \Delta t, y) dy - \int_a^b u(t, y) dy = \Delta t f(a) u(t, a) - \Delta t f(b) u(t, b).$$

The sign of  $f$  at points  $a$  and  $b$  determines whether cells enter or leave the region. For instance, if  $f(a) > 0$ , the first term of the right-hand side is positive, which translates the fact that cells enter the region  $[a, b]$  at point  $a$ . Similarly, if  $f(a) < 0$ , cells leave the region  $[a, b]$  at point  $a$ . On the other hand, if  $f(b) > 0$ , cells will leave the region  $[a, b]$  at point  $b$ , and if  $f(b) < 0$ , cells will enter the region  $[a, b]$  at point  $b$ . Intuitively, cells will then be transported towards the right when  $f$  is positive, and towards the left when  $f$  is negative. Concentration phenomena will happen at points where  $f$  is zero and has negative derivative: these points are asymptotically stable points for the ODE  $\dot{y} = f(y)$ . Figure 1 illustrates a toy situation in which  $f$  has a stable equilibrium point towards the middle of the phenotypic space, resulting in concentration of the cell density around this point (Fig. 1 left middle panel).

We also choose to take into account population growth, which includes cell division and cell death. Each cell of phenotype  $y$  divides at rate  $r(y)$  and its daughter cells are assumed to be given the same phenotype  $y$ . Hence, the quantity of new cells at time  $t + \Delta t$  with phenotype in  $[a, b]$  is given by  $\Delta t \int_a^b r(y) u(t, y) dy$ . Cells are also assumed to die at a rate  $d(y)\rho(t)$ , proportional to the total number of cells  $\rho(t) = \int_{\mathbb{R}} u(t, y) dy$ . The quantity of cells that died between the times  $t$  and  $t + \Delta t$  is then approximated by  $\Delta t \int_a^b d(y) \rho(t) u(t, y) dy$ . The term ' $r(y) - d(y)\rho(t)$ ' thus represents the net growth rate, where  $r(y)$  is the division rate, and  $d(y)\rho(t)$  the death rate. It is positive if cell division is faster than cell death, and negative in the opposite case. Overall, the evolution of the number of cells between times  $t$  and  $t + \Delta t$

**Fig. 1 | Growth, state regulation and stochastic state transition combine to give the cell density evolution in the proposed Cell Population Balance model.**

In each panel, the cell density  $u$  is represented at time  $t$  in black, and at a subsequent time  $t + \Delta t$  in colour. The number of cells with cell state in the interval  $[a, b]$  is represented by the hashed (black or coloured) area. Growth (which includes cell division and cell death) results in an upward or downward vertical shift of the cell density. In this toy example (top left panel), cell division takes place faster than cell death, so the cell density at time  $t + \Delta t$  is shifted upward compared to the cell density at time  $t$ . Consequently, the number of cells with cell state in  $[a, b]$  increases between  $t$  and  $t + \Delta t$ . State regulation results in a horizontal displacement of the cell density. In this example (center left panel), a concentration phenomenon is taking place towards the center of the state space. Here, the number of cells with cell state in  $[a, b]$  increases between  $t$  and  $t + \Delta t$ . Stochastic state transition leads to a mixing of the cell population. The evolution of the number of cells with cell state in  $[a, b]$  depends on the cell density at every other point of the state space. This results in a diversification of the population, that is in a flattening of the cell density curve (bottom left panel). In this example, the number of cells with cell state in  $[a, b]$  decreases between  $t$  and  $t + \Delta t$ . All three mechanisms combine in the cell population balance model to give the evolution of the cell density (right panel).



caused by the cell population growth is given by

$$\int_a^b u(t + \Delta t, y) dy - \int_a^b u(t, y) dy = \Delta t \int_a^b (r(y) - d(y)\rho(t)) u(t, y) dy.$$

Depending on the sign of the right-hand side, the growth mechanism will result in an upward or downward vertical shift of the curve representing the cell density (see Fig. 1, top left panel).

The third mechanism that we take into account is ‘stochastic cell-state transition’, that we will also refer to as cell-state transition. Here,  $M(y, z)$  represents the (infinitesimal) probability that a cell’s phenotype  $z$  changes to a phenotype  $y$ . Thus the number of new cells with phenotypes in the interval  $[a, b]$  at time  $t + \Delta t$  is computed as  $\Delta t \int_a^b (\int_{\mathbb{R}} M(y, z) u(t, z) dz) dy$ . Symmetrically, the number of cells whose phenotypes were in  $[a, b]$  at time  $t$  and who mutated between times  $t$  and  $t + \Delta t$  is computed by  $\Delta t \int_a^b (\int_{\mathbb{R}} M(z, y) dz) u(t, y) dy$ . This mechanism is fundamentally non-local, in that the evolution of the number of cells with phenotypes in the interval  $[a, b]$  depends on the whole cell population. Often, this mechanism results in a mixing of the population, that is in a flatter cell density (as illustrated in Fig. 1, bottom left panel).

Putting everything together, the variation of cells whose phenotype belongs to  $[a, b]$  during a time interval  $\Delta t$  is approximated by

$$\int_a^b u(t + \Delta t, y) dy - \int_a^b u(t, y) dy = \Delta t \underbrace{\int_a^b (r(y) - d(y)\rho(t)) u(t, y) dy}_{\text{growth}} \quad (2)$$

$$+ \underbrace{f(a)u(t, a) - f(b)u(t, b)}_{\text{advection}} \quad (3)$$

$$+ \underbrace{\int_a^b \left( \int_{\mathbb{R}} M(y, z) u(t, z) dz \right) dy - \int_a^b \left( \int_{\mathbb{R}} M(z, y) dz \right) u(t, y) dy}_{\text{cell-state transition}} \quad (4)$$

The combination of the three mechanisms is illustrated in Fig. 1 (right panel).

It is worth stressing that, all in all, we obtain a model where the interaction between cells happens solely at the level of the growth term. It is mediated by the non-local logistic term involving the total population size  $\rho(t)$ .

Taking the limit  $\Delta t$  going to zero in Equation (2), we obtain a PDE modelling the mechanisms of advection, growth and cell-state transition. Generalising to any dimension  $n$ , it can be written as

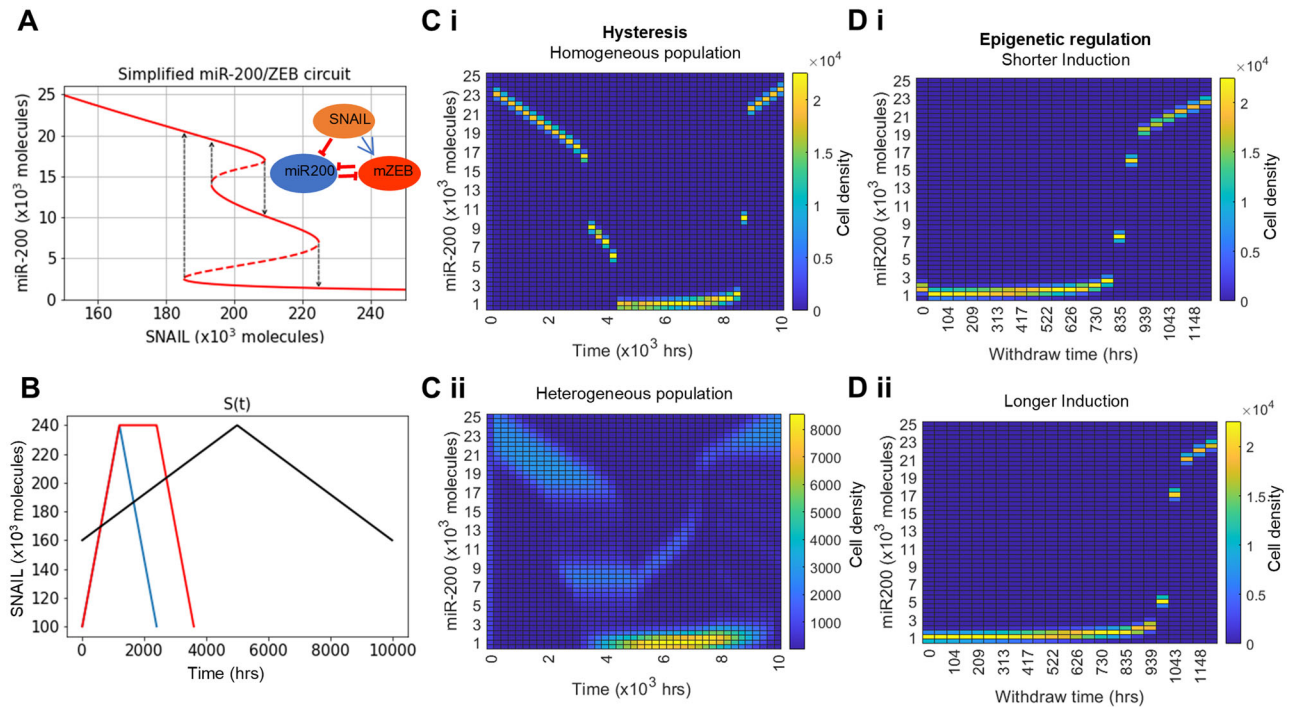
$$\begin{aligned} \partial_t u(t, y) + \nabla \cdot (f(y)u(t, y)) \\ = (r(y) - d(y)\rho(t))u(t, y) \\ + \int_{\mathbb{R}^n} M(y, z)u(t, z)dz - \int_{\mathbb{R}^n} M(z, y)dz u(t, y), \end{aligned} \quad (5)$$

and then complemented with an initial condition describing the density of cells at time 0.

### Incorporating EMT state regulation in the developed model captures hysteresis in EMT-MET cycle, and delayed MET after prolonged EMT induction

We first focused on the case where cell-state transition and growth in (5) are neglected. In this case, equation (5) reduces to the so-called *advection equation* (also often named *transport equation*), namely

$$\partial_t u(t, y) + \nabla \cdot (f(t, y)u(t, y)) = 0. \quad (6)$$



**Fig. 2 | Proposed cell population balance model reproduces hysteresis and epigenetic regulation of state transitions at the population levels.** **A** EMT gene regulatory network (inset) and the bifurcation of cell states resulting from the network dynamics with increasing input signal (SNAIL) levels, **B** Variation in the external input (SNAIL) levels with time to capture hysteresis and epigenetic regulation of cell states in **(C, D)**. **C** Hysteresis (non-symmetric trajectories) in cell state

transition during one cycle of EMT and MET by varying SNAIL levels (black curve) as shown in **(B)** while considering—(i) Homogeneous population and (ii) Heterogeneous population. **D** Dynamics of miR-200 levels in homogeneous population during withdrawal of SNAIL levels post short-term (i) and long-term (ii) EMT induction (blue and red curve in **B**, respectively). Growth of cells and state transition because of biochemical noise are neglected in these simulation results.

This equation is the exact PDE analogue of the ODE

$$\dot{y} = f(t, y). \quad (7)$$

In vitro and in silico studies on EMP have demonstrated two specific dynamic phenomena as cancer cells undergo one cycle of EMT and MET:

1. Asymmetry in EMT and MET trajectories (hysteretic behaviour) of cell states.
2. Delayed MET with an increasing duration of EMT inducer treatment due to epigenetic (chromatin-based) stabilisation of M and H states.

To properly define the advection function  $f$  underlying state regulation, we chose a minimal EMT regulatory network with canonical epithelial (microRNA-200 (miR-200)) and mesenchymal (mRNA ZEB) players that mutually inhibit each other via transcriptional and translational regulation. An EMT-inducing transcription factor SNAIL that activates ZEB and inhibits miR-200 represents the cumulative effect of upstream signalling pathways<sup>40</sup>. The corresponding ODE model is given by

$$\begin{cases} \dot{\mu}_{200} = g_{\mu_{200}} H_{Z, \mu_{200}}(Z) H_{S, \mu_{200}}(S) - g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) Q(\mu_{200}) - k_{\mu_{200}} \mu_{200} \\ \dot{Z} = g_Z g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) P(\mu_{200}) - k_Z Z \end{cases} \quad (8)$$

This ODE system is the modelling core for the advection term in all our simulations. It will progressively be made more complex in the corresponding section, or approximated by a simpler model to ease the computational burden when necessary (see next section). All the corresponding parameters are given in ‘Parameters for ODE (8) and (9)’.

We denote  $F = F(\mu_{200}, Z, S)$  the right-hand side of this ODE model, which accounts for interactions between a transcription factor ZEB (denoted  $Z$ ), and a micro-RNA miR-200 (denoted  $\mu_{200}$ ). The variable  $S$

represents a third molecule, SNAIL, which is seen in our case as an external signal characterising the extracellular environment. Note that this model falls into the framework introduced in equation (7), considering that  $y$  is the vector of miR-200 and ZEB concentrations  $y = (\mu_{200}, Z)$ , and taking  $f(t, y_1, y_2) = F(\mu_{200}, Z, S(t))$ .

The bifurcation diagram (Fig. 2A) depicts the different possible stable states, each characterised by a specific range of miR-200 levels (solid lines) for increasing levels of SNAIL, resulting from the network dynamics (8). As a cell undergoes EMT (i.e. SNAIL levels increase), it switches from high to intermediate to low levels of miR-200 which corresponds to E, H and M state respectively. However, during MET, the cell switches directly from low (M) to high (E) miR-200 levels, thus displaying hysteresis.

To reproduce the hysteretic behaviour of EMP, i.e. asymmetry in EMT and MET trajectories, we first simulated the advection equation (6) associated to ODE (8). In other words, we are considering (6) where  $y = (\mu_{200}, Z)$  is the two-dimensional vector of miR-200 and ZEB concentrations, the advection function is given by  $f(t, \mu_{200}, Z) = F(\mu_{200}, Z, S(t))$ , and  $t \mapsto S(t)$  is the piecewise-linear function which connects the points (0, 160K), (5000, 240K) and (10,000, 160K), as represented in Fig. 2B (black; here K represents 10<sup>3</sup> molecules).

We confirmed that the Cell Population Balance model developed here captures hysteresis, upon using SNAIL dynamics shown in black curve (Fig. 2B). For a homogeneous cell population (Fig. 2C, i), we see that the cells reside in three distinct miR-200 states (high, intermediate, and low) for time 0–5000 h (increasing SNAIL levels), but make a quick transition from low to high miR-200 levels for time 5000–10,000 h without spending much time in the intermediate state (decreasing SNAIL levels). The intermediate miR-200 levels seen during MET are to be understood as a sample timepoint where miR-200 levels are responding to changes in SNAIL levels before settling to their equilibrium (high) state.

In order to account for heterogeneity within the population, and more specifically for the fact that the signal  $S$  can be interpreted in a different way



by each cell, we incorporated  $S$  within the structure variable, which becomes  $y = (\mu_{200}, Z, S)$ . SNAIL level variation then impacts the advection term, which becomes  $f(t, y) = (F(y), f_S(S))$ , where  $f_S$  is the step function corresponding to the derivative of the function  $S$  introduced in the previous paragraph, i.e.  $f_S(S) = 40$  if  $S \in [0, 5000]$ , and  $f_S(S) = -40$  if  $S \in (5000, 10,000]$ . As initial condition, we considered a population homogeneously distributed in the molecules ZEB and miR-200, but with heterogeneous levels of SNAIL distributed according to a truncated Gaussian. In other words, we let  $u^0(\mu_{200}, Z, S) = \frac{1}{\sigma} G\left(\frac{S-S_0}{\sigma}\right) \mathbb{1}_{[50K, 250K]}(S) \mathbb{1}_{[0, 25K]}(\mu_{200}) \mathbb{1}_{[0, 1K]}(Z)$ , where  $S_0 = 160K$ ,  $\sigma = 20K$  and  $G$  denotes the Gaussian function, i.e.  $G(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ .

In this case, we obtain hysteresis as well (Fig. 2C, ii). Particularly, the cell distribution along ZEB and miR-200 axis during MET shows that the transient intermediate miR-200 peak in homogeneous population has turned into little dispersed transient peaks which were clearly distinct from the intermediate peak arising during EMT (Fig. S1). This observation corroborates with the experimental data on different partial states seen during EMT vs. during MET<sup>10</sup>. Also, in the heterogenous population case, some cells fail to complete a full EMT, rather undergo a partial transition and then return to an epithelial state upon reduced SNAIL levels (Fig. 2C, ii).

The range of SNAIL values for which E, H and M states are stable (Fig. 2A) can be altered by epigenetic (chromatin-based) changes that can happen during long-term EMT induction, leading to a delayed MET<sup>41,42</sup>. Therefore, to capture this phenomenon within our modelling framework, we incorporated the phenomenological formalism to account for epigenetic changes in EMT regulatory network (equation (9),<sup>43</sup>). In order to do so, we considered the ZEB's threshold for inhibiting miR-200, ' $Z_0$ ', to be a time-dependent variable such that the longer ZEB remains in the high states during EMT, the lower  $Z_0$  becomes. This lowered  $Z_0$  value enables ZEB's inhibition of miR-200 to continue even if ZEB level goes down when the external inducer is withdrawn<sup>42-44</sup>. To incorporate this phenomenological approach to model epigenetic regulation in the model, we correspondingly augmented the structure variable, which becomes  $y = (\mu_{200}, Z, Z_0, S)$ . The considered advection function is that associated to the ODE:

$$\begin{cases} \dot{\mu}_{200} &= g_{\mu_{200}} H_{Z, \mu_{200}}(Z_0, Z) H_{S, \mu_{200}}(S) - g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) Q(\mu_{200}) - k_{\mu_{200}} \mu_{200} \\ \dot{Z} &= g_Z g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) P(\mu_{200}) - k_Z Z \\ \dot{Z}_0 &= \frac{1}{\beta(t)} \left( Z_{\mu_{200}}^0 - Z_0 - \alpha Z \right). \end{cases} \quad (9)$$

All corresponding parameters are detailed in 'Parameters for ODE (8) and (9)'. Denoting  $F_e$  the right-hand side of this ODE, the corresponding PDE model (6) has advection function given by  $f_e(t, \mu_{200}, Z, Z_0, S) = F_e(\mu_{200}, Z, Z_0, S(t))$ . In a first simulation (Fig. 2D i),  $S$  is the piecewise-linear function which connects the points (0, 100K), (1200, 240K) and (2400, 100K), as represented in Fig. 2B (blue curve), and in a second simulation (Fig. 2D ii), the piecewise-linear function which connects the points (0, 100K), (1200, 240K) and (2400, 240K) and (3600, 100K) (red curve in Fig. 2B).

We considered two SNAIL levels dynamics to show the influence of epigenetic changes during EMT: Short-term induction (Fig. 2B blue curve), and Long-term induction (Fig. 2B red curve). We see that homogeneous cell population have longer recovery time for long-term EMT induction than for short-term induction, therefore, recapitulating the previous observation based on population's average cell analysis<sup>43</sup>.

### Incorporating E-M subpopulations' cell growth and spontaneous state transition in the Cell Population Balance model

We then moved on to considering the full model (5). In this case, precise models for the functions underlying the growth and cell-state transition terms are not as established (or even missing), making our approach more qualitative than quantitative.

When incorporating growth and cell-state transition into the model, computation times become prohibitive. We hence used dimension reduction by further simplifying the advection function: we considered the state

$y = (x, S)$  rather than  $(\mu_{200}, Z, S)$ , where  $x$  is roughly equivalent to  $\mu_{200}$ , as explained below.

This requires to choose a function  $f_r$  such that the dynamics of  $x$  for any  $S$  is given by  $\dot{x} = f_r(x, S)$ . Our main requirement in choosing this new function is to preserve the bifurcation diagram given in Fig. 2A, which means that, for a given value of  $S \in [150K, 250K]$ ,  $f_r(\cdot, S)$  has the same number of zeros as  $F(\cdot, S)$ , and are such that  $f_r(x, S) = 0$  if and only if there exists  $Z > 0$  such that  $F(x, Z, S) = 0$ .

Since infinitely many functions satisfy this property, one must make further assumptions in order to select a suitable one. For simplicity and to avoid overfitting, we assumed that, for any  $S$ ,  $f_r(\cdot, S)$  is piecewise linear (with one root per interval where it is linear) and that the rate of change is constant on each interval. Under these constraints, the function is defined up to a multiplicative constant which is chosen by minimising a suitably defined criterion, see 'Reduction of the advection function'.

Figure S2 A shows the shape of the function  $x \mapsto f_r(x, S)$  underlying the reduced ODE for different values of  $S$ , and highlights the relative stability of possible cell-states for increasing levels of input  $S$ .

After reduction, and considering now growth and cell-state transition, the full PDE model incorporating growth and stochastic state transition is of the form (5) in dimension  $n = 2$ , namely

$$\begin{cases} \partial_t u(t, y) + \nabla \cdot (f(y) u(t, y)) = (r(y) - d(y) \rho(t)) u(t, y) \\ + \int_{\mathbb{R}^2} M(y, z) u(t, z) dz - \int_{\mathbb{R}^2} M(z, y) dz u(t, y) \\ \rho(t) = \int_{\mathbb{R}^2} u(t, y) dy \end{cases}, \quad (10)$$

where:

- The structure variable  $y \in \mathbb{R}^2$  is  $y = (x, S)$ .
- The advection function  $f$  is given by  $f(y) = f(x, S) = (f_r(x, S), f_S(S))$ , where

$$f_S(S) = \delta \left( 1 - \frac{S}{S_0} \right), \quad (11)$$

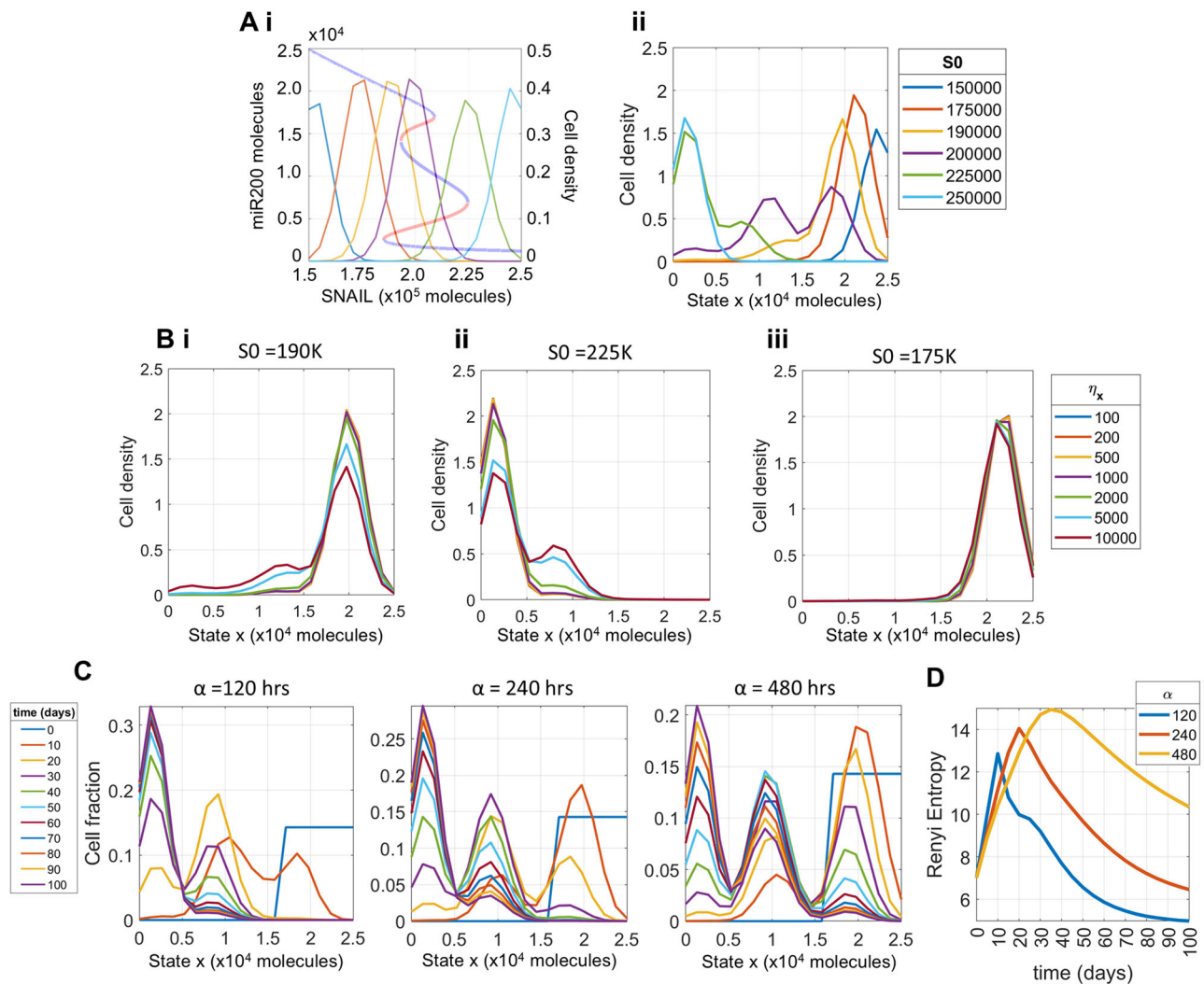
with  $S_0 \in [150K, 250K]$  corresponding to the mean of the SNAIL distribution,  $\delta := \frac{S_0 \ln(2)}{\alpha}$ , and  $\alpha > 0$  representing the characteristic time of convergence of SNAIL to the mean  $S_0$ .

- Within the term  $(r(y) - d(y) \rho(t)) u(t, y)$  representing the net growth of cells of state  $y$ , the death rate is always considered to be independent of the cell state (i.e.  $d(y) = d \equiv 1.82 \times 10^{-7}$  cell/h, as in ref. 32).
- The 'mutation' function  $M$  is taken to be  $M(y, z) = r(z) P(y - z) \mathbb{1}_{[0, 25K] \times [150K, 250K]}(y)$ . This choice is motivated by the assumption that transitions occur at cell division. We note that other modelling assumptions might lead to the same choice. The presence of indicator functions ensures that the support of the solution remains in the rectangle  $[0, 25K] \times [150K, 250K]$  at all times. Here,  $P(y) = P(x, S) := \frac{1}{\eta_x \eta_S} G\left(\frac{x}{\eta_x}\right) G\left(\frac{S}{\eta_S}\right)$ , where  $G$  is the Gaussian function. Hence, a cell defined by state  $x, S$  gives birth to daughter cells defined by state  $x', S'$ . Our models captures how far  $x'$  and  $S'$  are from  $x, S$  through standard deviations  $\eta_x$  and  $\eta_S$ , respectively.

To solve (10), we make use of a particle method adapted to this problem, as introduced in ref. 39. The details of this numerical method are given in 'Numerical method'.

### Biochemical reaction noise, coupled with regulatory cell-state dynamics, shapes heterogeneity pattern of the cell population

To evaluate how stochastic processes influence the population distribution of E, M and hybrid H states, we considered the scenario where the cell division rate  $r(y)$  is independent of the cell-state  $y$  (i.e. assuming EMT does not impact cell cycle). Because the levels of EMT-inducing signal SNAIL are also evolving due to biochemical reaction noise, its levels are distributed in the population around the mean environmental characteristics  $S_0$  (Fig. 3A, i). We first established similarity between the dynamics of the system with



**Fig. 3 | Simulations of the PDE model (10) with reduced function  $f_r$ .** Empirically reduced single cell state variable  $x$ , showing that the reduced system captures the full EMT network dynamical response; and the combined influence of biochemical noise with input signal dynamical response in modulating the phenotypic heterogeneity of the cell population. **Ai** Bifurcation diagram showing stable (blue) and unstable (red) levels of miR-200 (y axis on left) for increasing SNAIL levels; and distribution of SNAIL levels among the cells (y axis on right) for different mean characteristics levels  $S_0$ . **Aii** Distribution of cells along the state  $x$  of reduced EMT system for different distributions of SNAIL shown in (Ai) after a simulation time of 100 days starting from epithelial cells. **B** Changes in population's phenotypic distribution with

increasing levels of epigenetic noise  $\eta_x$  for SNAIL distributions with mean  $S_0 = 190K$  (Bi) and  $225K$  (Bii). For  $S_0 = 175K$  (Biii), the population remains invariant to increasing noise. **C** Temporal changes in cell state distribution of the population for decreasing values of input signal SNAIL's perturbation recovery rate (increasing values of the characteristic time ' $\alpha$ '). **D** Time evolution of population's heterogeneity (measured by differential entropy) for different  $\alpha$  parameter values. Parameters used to generate plots, unless stated otherwise, are  $\alpha = 120$  h,  $\eta_x = 1000$ , ini pop Epi (see Fig. 4A, i), time = 100 days,  $S_0 = 225K$  molecules, and division rate  $r$  is constant across phenotypes, given by  $r = 0.0182/h$ .

two variables (full EMT) and that of the reduced system by comparing the distribution of miR-200 levels for a given SNAIL distribution characteristic with the cell population distribution along the state ' $x$ ' (Fig. 3A and Fig. S2B).

For example, with a SNAIL distribution of mean value ( $S_0$ ) of 200 K molecules, the bifurcation diagram depicts the possibility of the cell population to be distributed in all three states (Fig. 3Ai). We find that the asymptotic distribution of the state variable  $x$  from reduced system dynamics exhibits tri-modality, showing the co-existence of all three states, irrespective of the initial population condition (Fig. 3A, ii—initial population: all cells as epithelial, Fig. S2B—initial population: all cells as hybrid or mesenchymal). Similarly for a SNAIL distribution with mean  $S_0 \in \{190K, 225K, 150K, 250K\}$  molecules, we observe respective combinations of phenotypes as in the bifurcation diagram of the full EMT network - co-existence of E and H (bimodal), co-existence of H and M (bimodal), epithelial (unimodal) and mesenchymal (unimodal) (Fig. 3A), thus providing further evidence of faithfully representing EMT dynamics through the sole variable  $x$ .

For multi-modal distributions of the state variable  $x$ , the exact phenotypic composition depends on relative stability of the multiple co-existing phenotypes. For example, we see a reduced share of H (intermediate  $x$  levels) cells for distribution of SNAIL levels that overlap significantly with those that have mean SNAIL levels corresponding to monostable E or monostable M regions ( $S_0 = 190K, 225K$  molecules respectively), especially for a reduced standard deviation  $\eta_x$  of stochastic cell-state transition in state  $x$  (Fig. 3Bi, ii). Similarly, for distribution of SNAIL levels with mean  $S_0 = 175K$  molecules, although both E and M states co-exist (Fig. 3A, i), the relative stability of the E state is much greater than that of the M state, thus disallowing cells to make transition to the M state even at higher levels of  $\eta_x$  (Fig. 3B, iii).

As mentioned previously, the distribution of SNAIL levels and correspondingly that of the cell state  $x$  in a population can be attributed to stochasticity in biochemical reactions. Another type of perturbation in cellular variables (here,  $x$  and SNAIL) can arise when certain

subpopulations are isolated and re-cultured independently. For instance, when the E, M and H prostate cancer subpopulations are segregated, they exhibit very different distributions after 2 weeks<sup>12</sup>. Similarly, the segregated EpCAM-high and EpCAM-low subpopulations in breast cancer have varied recovery dynamics<sup>45</sup>. Thus, in case of either internal (stochastic cell-state transition) or external (micro-environmental factors) perturbations, the rate at which the cellular variables recover towards the characteristic distribution can differ even though they eventually converge to the same equilibrium distribution. Thus, we next modulated the rate of recovery of SNAIL levels to mimic the scenario of extrinsic perturbation to the cell population by isolating distinct subpopulations and simulating (re-culturing) them independently.

The rate of recovery to perturbations in SNAIL levels is inversely correlated to the parameter  $\alpha$  in our model, which captures the characteristic time of convergence of SNAIL levels to its equilibrium  $S_0$ <sup>43,46</sup>. For an extrinsic perturbation (e.g. enriched epithelial cells from a M cell majority population), we see that the time evolution of the population distribution slows down with increasing values of  $\alpha$  (Fig. 3C). Furthermore, the slowed down dynamics increases cellular heterogeneity by causing the population to be distributed in all three states for a considerable interval of time, as quantified by differential entropy (Fig. 3D). In the example shown, population heterogeneity first increases as the population shifts from a majority of epithelial cells to being more uniformly distributed among the three phenotypes in the intermediate time points, and then decreases as the population becomes mainly composed of M cells. Similar observations can be made for other combinations of initial condition and mean  $S_0$  values of SNAIL distribution (Figs. S3, S4).

Overall, the interplay between deterministic and stochastic dynamics of cellular biomolecules shapes the population heterogeneity. This is done both by distributing cells in all plausible states permitted by the underlying regulatory network dynamics, and by slowing down the kinetics of cells towards equilibrium when perturbed by the external signal S.

### Difference in phenotypic division rates reduces E-M heterogeneity, which could be recovered by increasing biochemical noise levels

We performed model simulations to see how an interplay between deterministic and stochastic effects on the cell state  $x$ , and input SNAIL, can impact the E-M heterogeneity patterns. So far, we considered all three phenotypic states (E, M and H) to divide at equal rates. To observe the additional influence of division rate differences on E-M heterogeneity we considered three possible scenarios—(a) case ‘r1’: all three phenotypes divide at the same rate, (b) case ‘r2’: both E and H cells divide at equal rates, while M cells divide at half the rate of E cells; and (c) case ‘r3’: both H and M divide at equal rates, which is half the rate of division of E cells. In practice, this means considering three different possible piecewise-constant functions  $r$ , as illustrated in Fig. 4A, ii. Across all these cases, the E state divides at either an equal or a faster rate than H and/or M cells. This constraint recapitulates the current experimental understanding that EMT may suppress cell cycle to varying extents, thus reducing the division rate of H and/or M cells<sup>35,47</sup>.

First, we investigated how the phenotypic composition of the population changes with different division rate scenarios (Fig. 4, Fig. S5). For uniformly distributed cells in all three E (epi), H (hyb), and M (mes) states (Fig. 4A, i) and SNAIL distribution with mean  $S_0 = 190$  K or 200 K molecules that predominantly enables an E state with/without H state, the reduced growth of M cells has very slight effects on phenotypic composition over the time course, as expected (Fig. 4B, i). The initial peak in hybrid cell fraction for the ‘r2’ growth scenario is the combined effect of M to H state transition and a relatively higher division rate of H cells. However, when H cells also have reduced proliferation (growth scenario ‘r3’), we see a lasting change in the phenotypic composition as E cells become dominant because of higher division frequency (Fig. 4B, i— $S_0 = 200$ K). For the input SNAIL distribution with a mean value  $S_0$  of 225 K molecules that majorly supports H and M phenotypes, in the case ‘r2’, the growth advantage provided to H

cells enables their dominance in the population on the long run, when compared to the growth scenarios of ‘r1’ or ‘r3’ where both H and M cells proliferate at equal rates (Fig. 4Bi,  $S_0 = 225$ K). The initial peaks in H fractions are combined effects of E to H transitions with either growth similarity or advantage of H cells over M cells.

The overall change in phenotypic composition can be calculated using differential entropy as a heterogeneity score (Fig. 4B, ii). Although growth scenarios ‘r1’ and ‘r2’ have the same phenotypic composition and an equal heterogeneity score eventually, the growth scenario ‘r1’ shows a much smoother change in heterogeneity values from the initial levels because of all three phenotypes being equally proliferative. Next, we looked at the effects of increasing level of epigenetic noise level  $\eta_x$  in cell state  $x$  on phenotypic composition laid down by division rate differences (Fig. S6A–D). For  $S_0 = 190$ K and 200K, where the H state is less dominant than the E state (Fig. 4B), increasing the noise levels (from  $\eta_x = 1000$  to  $\eta_x = 5000$ ) in state  $x$  causes more cell-state transitions, raising the frequency of the H phenotype in population for all growth scenarios (Fig. S6B). However, as for  $S_0 = 225$ K and growth scenario ‘r2’ where the hybrid phenotype is the dominant state, increasing the noise  $\eta_x$  level (from  $\eta_x = 1000$  to  $\eta_x = 5000$ ) raises M fractions in the population even though M cells are dividing slowly (Fig. 4B, i; Fig. S6B, C). Overall, we observe an increase in population heterogeneity with higher noise levels in state variable ‘ $x$ ’, irrespective of the growth scenario (Fig. 4Bii, Fig. S6D).

After observing changes in phenotypic composition for different growth scenarios, we moved on to see how total cell population grows for combinations of initial conditions and growth scenarios. We considered six different conditions—isolated E, H, and M population, uniform mixture of either E and M or E, H and M cells, and uniform distribution of cells in all possible cell states  $x$  and input SNAIL levels.

When all the phenotypic states are dividing at equal rates, the total number of cells does not vary across different initial conditions. However, with M dividing slower than E and H cells (growth scenario ‘r2’), we see that an initially mesenchymal population has slower population growth compared to other initial conditions. Similarly, with both H and M cells dividing slower than E (growth scenario ‘r3’), initial conditions having component of either H or M cells divide slower than isolated (pure) E cells. Also, as transition from H to E is much more probable than M to E transitions, presence of hybrid cells in the population increases the overall division rate (Fig. 4C compare ‘hyb vs mes’, and ‘epi mes’ vs ‘epi hyb mes’ initial conditions for  $S_0 = 200$  K, ‘r3’ growth scenario). Further, increasing the level of cell transition stochasticity (from  $\eta_x = 1000$  to  $\eta_x = 5000$ ) causes more state transitions rendering lesser variability in the growth dynamics by quickly equalising the effect of differences in the initial conditions (Fig. S6E).

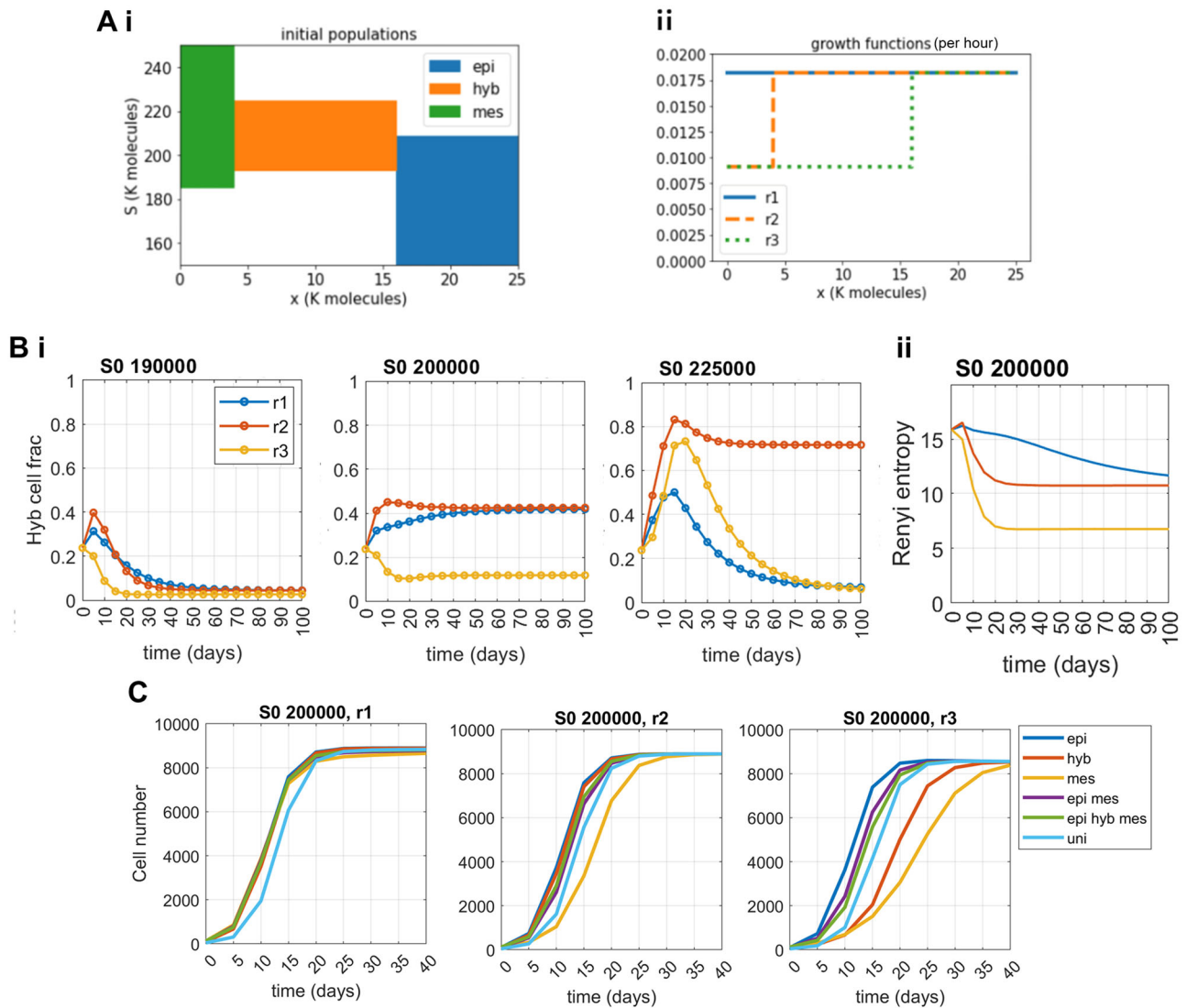
### Addition of resource competition among subpopulations enables bistability in phenotypic composition and bi-phasic growth of total cell number

Although there was limitation to growth to the cell population so far via a logistic growth term, all subpopulations were considered to present equal resource competition to each other. More specifically, the death rate  $d(y)p(t)$  in Equation (5) is considered to be proportional to the total population size  $\rho(t)$ , which supposes that all three subpopulations consume resources in the same amount. Here, we investigated how unequal resource competition among E, M and H affects the dynamic and asymptotic phenotypic composition of the cell population.

For this purpose, we developed a generalisation of Equation (5) with a more general growth term. For simplicity, we assumed that the SNAIL level is constant and equal to  $S_0 = 200$  K for all cells at all time. This generalised equation can then be written as

$$\begin{cases} \partial_t u(t, x) + \nabla \cdot (f(x)u(t, x)) = (r(x) - \rho_\kappa(t, x))u(t, x) \\ + \int_{\mathbb{R}} M(x, z)u(t, z)dz - \int_{\mathbb{R}} M(z, x)u(t, x)dz \\ \rho_\kappa(t, x) = \int_{\mathbb{R}} \kappa(x, z)u(t, z)dz \end{cases}, \quad (12)$$





**Fig. 4 | Effect of division rate differences among E, H and M phenotypes on population's heterogeneity and overall growth dynamics.** **Ai** Support of the different initial conditions: In each case, the initial condition is uniformly distributed on its support, and such that the total initial population equals 100 cells. **Aii** Profile of different growth functions, expressed per hour; 'r1': All three phenotypes divide at same rate; 'r2': E and H divide at equal rates, while M divide at half the rate of E cells; and 'r3': Both H and M divide at equal but half the rate of E cells. **Bi** Temporal changes in hybrid cell fraction in the population for different growth scenarios among phenotypes. **Bii** Time evolution of population's heterogeneity (measured by

differential entropy). **C** Population growth dynamics for different combinations of growth scenarios and initial condition; 'epi hyb mes' corresponds to an initial condition uniformly distributed on the three coloured domains of A1, and 'uni' to a uniform population on the whole rectangle  $[0, 25 \text{ K}] \times [150 \text{ K}, 250 \text{ K}]$ . For (B), the initial condition is uniformly distributed in all states. The input SNAIL mean ( $S_0$ ) levels used are mentioned for all the individual plots. Other parameters used to generate plots are  $\alpha = 120 \text{ h}$ ,  $\eta_x = 1000$ , and the division rate ( $r$ ) of the epithelial phenotype is  $0.0182/\text{h}$ .

where

- The advection function  $f$  is given by  $f(x) = f_r(x, S_0)$ .
- The 'mutation' function  $M$  is taken to be  $M(x, z) = r(x)P(x - z)\mathbb{1}_{[0, 25\text{K}]}(x)$ , with  $P(x) = P(x) := \frac{1}{\eta_x}G(\frac{x}{\eta_x})$ , where  $G$  is the Gaussian function.
- The death rate  $\rho_\kappa(t, x)$  is now computed as the integral of the cell density  $u(t, z)$  weighted by a kernel  $\kappa(x, z)$ .

The choice of  $\kappa$  allows to include competition for resources which depends on the cells' states. For instance,  $\kappa(x, z) > \kappa(z, x)$  models the fact that whenever cells with states  $x$  and  $z$  compete for resources, a cell with state  $z$  consumes more resources than a cell with state  $x$ . In the case where  $\kappa(x, z) = d(x)$ , we recover equation (5), with  $\rho_\kappa(t, x) = d(x)p(t)$ .

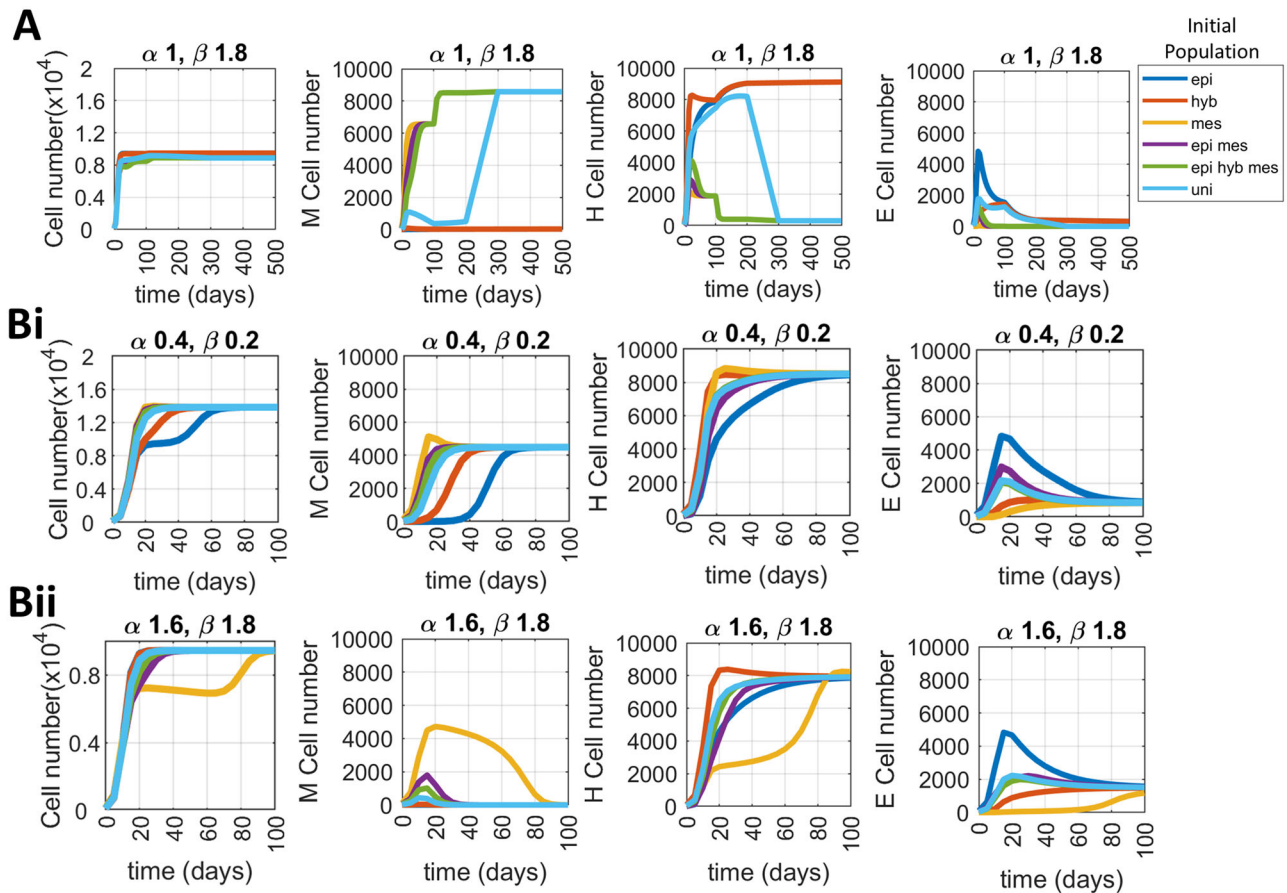
For simplicity, in the simulations of this section, we chose  $\kappa$  of the shape

$$\begin{aligned} \kappa(x, z) = & d_m(\alpha_{mm}\mathbb{1}_M(z) + \alpha_{hm}\mathbb{1}_H(z) + \alpha_{em}\mathbb{1}_E(z))\mathbb{1}_M(x) \\ & + d_h(\alpha_{mh}\mathbb{1}_M(z) + \alpha_{hh}\mathbb{1}_H(z) + \alpha_{eh}\mathbb{1}_E(z))\mathbb{1}_H(x) \\ & + d_e(\alpha_{me}\mathbb{1}_M(z) + \alpha_{he}\mathbb{1}_H(z) + \alpha_{ee}\mathbb{1}_E(z))\mathbb{1}_E(x), \end{aligned}$$

where  $\mathbb{1}_E(x) = 1$  if  $x \in E = [16\text{K}, 25\text{K}]$  (epithelial cells) and 0 otherwise,  $\mathbb{1}_H(x) = 1$  if  $x \in H = [4\text{K}, 16\text{K}]$  and 0 otherwise, and  $\mathbb{1}_M(x) = 1$  if  $x \in M = [0, 4\text{K}]$  and 0 otherwise.

This choice of  $\kappa$  is to be linked with the Lotka-Volterra resource competition model: indeed, in the absence of state transition and advection, and by taking  $r(x) = r_m\mathbb{1}_M(x) + r_h\mathbb{1}_H(x) + r_e\mathbb{1}_E(x)$ , one easily checks that





**Fig. 5 | Resource competition between E, M and H states enables multi-stability in phenotypic composition, and biphasic growth pattern in total cell counts.** Model (12) is used to model resource competition between E, M and H cells. The self-interaction parameters of the resource competition for all three phenotypes are set to 1 and the E and H cells were assumed to be identical in terms of resource consumption. Therefore, their cross-interaction parameters is set to unity. Parameter  $\alpha$  denotes resource competition from E and H to M cells and  $\beta$  denotes resource competition from M to E and H cells. A two distinct steady-state phenotypic compositions of M and H hybrid arise when initial population contains low and moderate-to-high proportion of the M cells. Multi-stability in phenotypic composition occurs in small region of parameters  $\alpha$  and  $\beta$  controlling resource competition (Figs. S8 and S9). **B** Biphasic growth pattern in total cell number. Biphasic growth is

defined as presence of both quasi and asymptotic steady states at intermediate and long-term time instances, respectively. **Bi** the resource competition among E, M and H cells is small (both  $\alpha$  and  $\beta < 1$ ) and the initial cell population is E dominant ('epi' starting population), and **Bii** the resource competition among E, M and H cells is large (both  $\alpha$  and  $\beta > 1$ ) and the initial cell population is M dominant ('mes' starting population). The input SNAIL mean ( $S_0$ ) = 200 K,  $\eta_x = 2000$ , and the division rate ( $r$ ) of the epithelial phenotype is 0.0182/h. Initial populations 'epi', 'hyb' and 'mes' correspond to 100 cells uniformly distributed on E, H and M respectively, 'epi mes' to a population of 50 cells on E and 50 cells on M, 'epi hyb mes' to a population of 33 cells on E, 33 cells on H and 33 cells on M, and uni to a population of 100 cells uniformly distributed on the whole  $[0, 25K] = E \cup H \cup M$ .

the epithelial, hybrid and mesenchymal populations, given respectively by

$$e(t) := \int_E u(t, z) dz, \quad h(t) := \int_H u(t, z) dz, \quad m(t) := \int_M u(t, z) dz$$

satisfy the ODE model

$$\begin{cases} \dot{m}(t) = r_m m(t) - d_m (\alpha_{mm} m(t) + \alpha_{hm} h(t) + \alpha_{em} e(t)) m(t) \\ \dot{h}(t) = r_h h(t) - d_h (\alpha_{mh} m(t) + \alpha_{hh} h(t) + \alpha_{eh} e(t)) h(t) \\ \dot{e}(t) = r_e e(t) - d_e (\alpha_{me} e(t) + \alpha_{he} h(t) + \alpha_{ee} e(t)) e(t) \end{cases}$$

In our simulations, we fixed  $r(x) = 0.0182 \text{ h}^{-1}$  (corresponding to 38 h of doubling time),  $d_m = d_h = d_e = 1.82 \times 10^{-7} \text{ /hr/cell}$  (assuming per cell death is  $10^4$  times slower than birth),  $\alpha_{mm} = \alpha_{hh} = \alpha_{ee} = \alpha_{eh} = \alpha_{he} = 1$ , and we assumed that  $\alpha_{me} = \alpha_{mh} =: \alpha$  and  $\alpha_{em} = \alpha_{hm} =: \beta$ . We ran simulations for different values of  $\alpha$  and  $\beta$ , and different initial distributions. Here, the assumption is that E and H cells behave identically in terms of resource consumption and in how they compete with M cells.

We observe that for certain values of  $\alpha$  and  $\beta$ , particularly  $\alpha \geq 1$  and  $\beta > 1$ , the phenotypic composition of the population has two asymptotic states, which can be reached depending on the initial distribution of the population (Fig. 5A, Figs. S7Ai, S8, S9). For example, the initial population containing more than 30% of M cells converges to high M cells fraction/numbers in the population, otherwise the asymptotic population is found to be dominated by H cells. Further, the bistable phenotypic composition remains intact even if the rates of cell-state transition are significantly reduced to minimal levels ( $\eta_x = 1$ , Fig. S7Aii). The range of competition coefficients ( $\alpha$  and  $\beta$ ) for which we have bistability in phenotypic composition may change when either the relative stability of E, M and H cells or their cell-state transition rates are varied (Fig. S10).

For certain initial population structures, we also observe biphasic growth in the total cell number, i.e. the presence of both quasi (apparent) as well as asymptotic steady states at the intermediate and long-term time instances, respectively (E 'epi' starting population in Fig. 5Bi and M 'mes' starting population in Fig. 5Bii). Since E and H cells present equal resource competition to each other, the total cells arising from starting E (epi) population saturates to the carrying capacity (10,000 cells) until the time population is devoid of M cells (40 days). Once M cells arise in the

population due to phenotypic transitions, the total population gets boosted as the resource competition is low ( $\alpha < 1$  and  $\beta < 1$ ). Thus, we see a biphasic growth pattern when starting with E population. Now, as the biphasic growth depends on the moment when the M cells arise in the population, reduced rate of cell-state transition ( $\eta_x = 1$ ) further extends the duration of time spent in quasi steady state of low total cell counts (Fig. S7Bi). We also see biphasic response when there is high resource competition among E(or H) and M cells ( $\alpha > 1$  and  $\beta > 1$ ) and M ('mes') starting population (Fig. S7Bii). Here, before the M cells reach the capacity of the population (10,000 cells), the high cell-state transition rates give rise to H cells within a short period of time, leading to significant resource competition and slowed down growth of both M and H cells. However, increasing population size of H cells and occurrence of E cells at later timepoints heavily lead to suppression of the M cells population by 80 days of simulation time. With reduced cell state transition, M cells gain dominance of the population before enough H cells appear in the population, and therefore, M cells suppress H and E cells number for all simulation time leading to loss of biphasic growth (Fig. S7Bii). When there is asymmetry in resource competition among E(or H) and M cells ( $\alpha < 1$  and  $\beta > 1$  or  $\alpha > 1$  and  $\beta < 1$ ), no significant patterns of bistability of phenotypic composition and biphasic total cell responses are seen (Figs. S8, S9 and S11).

## Discussion

Several intracellular and intercellular regulatory and stochastic processes shape heterogeneity within a population. While several intracellular processes, such as non-linear transcriptional regulation, chromatin-based epigenetic regulation and microRNA-mRNA binding and complex degradation either individually or in combination can result into the existence of distinct gene expression states, the stochastic intracellular processes such as transcriptional bursting, asymmetric cell division, and cell-to-cell communication lead the cells to switch from one gene expression pattern to the other. Experimental data capturing temporal changes in population level heterogeneity while profiling cells for transcriptome and epigenome is only recent in the context of EMT. However, significant efforts have been made in multi-scale mathematical modelling of a cell population that is growing, dividing and changing its phenotypic distribution with intracellular state dynamics based on one or more regulatory processes. Here, we contribute to this rich multi-scale population modelling literature by developing a framework allowing us to study how E-M population heterogeneity is regulated by—(1) the regulatory and stochastic intracellular processes, (2) heterogeneity of growth rates, and (3) resource competition among distinct subpopulations.

Our analysis is based on a minimalistic three node EMT regulatory network with a characteristic phenotypic landscape (Fig. 2A)<sup>40</sup>. Our choice of a sufficiently simple network was based on the following criteria—(a) making the analysis computationally tractable, and (b) integrating multiple processes together—cell division, cell death, cell-state transition and intracellular regulatory dynamics. However, many more EMT regulatory networks with increasing complexity (nodes and types of interactions) have been identified over the past decade<sup>13,24</sup>. The identification and addition of new players that regulate EMT modify the relative stability of epithelial (E), hybrid (H) and mesenchymal (M) cell states, and thus, will change the phenotypic composition of the population at steady state for a given input signal and noise level in our study<sup>48–50</sup>. Nonetheless, the advection term in the developed PDE model can be easily updated with other regulatory networks to observe the resulting changes in the population composition.

We find that the existence of E-M heterogeneity in the cell population is a consequence of noisy biochemical process and constitutive signalling set at levels where the regulatory network operates in mono-stable or multi-stable regime (Fig. 3). Such necessity of both noise and multi-stability for getting a heterogeneous population may explain why certain breast cancer cell lines have homogeneous composition of E or M cells and others cell-lines are heterogeneous<sup>15,30</sup>. Further, the heterogeneity imparted by biochemical reaction noise gets suppressed because of growth rates differences among E-M subpopulations (Fig. 4B). On the other hand, lower biochemical noise

can delay the rise in heterogeneity and growth of the total population if the phenotype composition of the population has a majority of slow growing cells to begin with (Fig. 4C)<sup>45</sup>. Such slow growth of cell population on reduced cell-state transitions may explain why stable mesenchymal cells are unable to form large tumours while plastic mesenchymal cells can in-vivo<sup>51</sup>.

The minimalistic EMT model considered here does not have a regulatory term for the input signal, so we assumed that the input has negative feedback onto itself and that its level in the population is distributed around the mean  $S_0$ . The parameters  $\alpha$  and  $S_0$  in input dynamics of S (Equation (11)) can be considered as the inverse of the birth rate and ratio of birth to death rates of S molecules, respectively. Therefore, the degradation and/or de-novo synthesis rates encoded in parameters  $\alpha$  and  $S_0$  set the rate at which equilibrium phenotypic distribution is attained when individual subpopulations are isolated (Fig. 3). Such role of biomolecule turnover rates in cell-state switching has also been reported in other in-vitro and in-silico studies. For example, while cells possessing a transgene with long half-life of mRNA and protein were able to pass on the transgene expression through multiple cell divisions, the transgene expression in cells with short mRNA and protein half-life could not be transmitted for more than a few generations<sup>52</sup>. Similarly, an in-silico study showed how switch in the expression of a gene can be hindered by long half-life of its mRNA molecule<sup>53</sup>. In the context of E-M plasticity, it was observed that different sets of molecular markers take varying time to revert to the basal expression levels post EMT induction<sup>54</sup>. Thus, the timescales of spontaneous cell-state switching are determined by the molecular markers used to define the cell states.

By incorporating resource competition between E, H and M subpopulations, we either saw biphasic growth in total cell number or initial condition dependent asymptotic subpopulations levels (Fig. 5)<sup>55,56</sup>. Such density-dependent growth interactions have been thoroughly studied in the dynamics of sensitive and resistant cancer cells and used to retrospectively explain in vivo/clinical data. However, in addition to resource competition, various other factors, such as (1) juxtacrine and paracrine signalling among cancer cell subpopulations, and (2) interactions of tumour cells with extracellular matrix (ECM) and/or other stromal cells) can alter cell growth rates and influence overall population structure<sup>9,57</sup>. For example, cells that undergo EMT can secrete LOXL2 which increases collagen crosslinking in ECM, and ECM density and stiffness can induce EMT<sup>58</sup>. The developed Cell Population Balance model can be extended to capture the cellular interactions mentioned above and is considered the future directions of this study.

Mathematical models that capture dynamics of population-level heterogeneity resulting from regulatory and population dynamics of cells include Agent-Based models and Cell Population Balance models. Among the advantages of Agent-Based models are their ability to capture the evolution of heterogeneity with details at single cell levels. However, simulating Agent-Based models leads to prohibitive computational costs for a large number of cells. On the other hand, Cell Population Balance models capture the evolution of cell densities in a heterogeneous population, therefore inherently assuming a large population size. Further, the cell density computed in Cell Population Balance models can be directly mapped to output of flow cytometry experiments conducted at multiple timepoints for a population. Thus, over the past two decades, Cell Population Balance models have been used to combine complex regulatory phenomena, e.g. positive feedback loops<sup>59</sup>, caspase activation cascade during programmed cell death<sup>60</sup>, cell-to-cell communication<sup>61</sup>, and two mutually inhibiting nodes<sup>62</sup>, with stochastic processes like asymmetric cell division and stochastic biochemical reactions for phenotype-structured, and age-structured cell populations<sup>61,63,64</sup>.

In our study, the simultaneous consideration of cell division, death, and state transition for the full EMT network lead to significant increase in the computational cost. Therefore, we reduced the existing model with two variables and one input signal to a model with a single variable and one input signal. The parameters of the reduced one-dimensional state derivative function (Fig. S2A) were set by minimising the error of its resulting dynamics with that of the full EMT model.

Our approximation and parameterisation of the state derivative function are in line with efforts over the last decade to use experimental flow cytometry and cell counts data to either parameterise the cell population balance model or optimally choose the functional form of state dynamics along with other parameters that fit the data well<sup>60,63,65</sup>. Notably, these studies use Maximum Likelihood approaches to parameterise the system since their aim is to fit data rather than approximate an already existing but more complex model.

To model stochasticity in the cell's state resulting from, for example, transcriptional bursting or asymmetric cell division, we have generalised intracellular noise by considering a general transition kernel, which we then assumed to be normally distributed around the current cell state by focusing on the case where stochasticity is mediated by cell division. We note that some literature has shown the individual contribution of each of the stochastic cellular processes on cellular heterogeneity<sup>59,61,66</sup>, and the model framework developed here can be easily used to account for other stochastic factors.

Our framework allows for the addition of a diffusion term, as has also been done in the literature. Indeed, a second-order diffusion term is the PDE counterpart of adding a Brownian motion to the considered ODE as in ref. 43. The corresponding diffusion term can be recovered by a suitable scaling of our integral transition term as explained in ref. 67, which we have not done in the present work. Additionally, assuming the noise to be normally distributed gave us a handle on its extent (standard deviation), and therefore enabled us to comment on the changes in population heterogeneity with increasing levels of noise (Fig. 3B).

The main equation studied in this article (10) is in line with Cell Population Balance models usually employed for heterogeneous cell populations<sup>62,64,68,69</sup>. Nevertheless, we have chosen a rather common logistic shape for the growth term, which writes  $(r(y) - d(y)\rho(t))u(t, y)$ , with  $\rho(t)$  the total number of cells at time  $t$ , rather than a simple linear term as it is often done. From a biological point of view, this choice reflects the capacity for the cell population to self-regulate its growth due to density constraints. From the mathematical perspective, it guarantees that the size of the population does not blow up, i.e. remains bounded, and allows one to study how the population evolves in larger times.

Regarding the numerical implementation, we opted for a scheme from the family of particle methods: these schemes are indeed known to be well-suited in the context of PDEs with advection and 'non-local' terms, that is terms that involve the density of the cell population at all points. In our case, the selection term and the state transition term are both non-local terms, since the former depends on the population size  $\rho$ , and the latter is a convolution with the so-called 'mutation function'  $M$ .

Convergence of particle methods has been proved under conditions that are satisfied in our setting<sup>39</sup>. Compared with other methods used for the same type of problems, such as finite element methods<sup>70</sup> or finite difference methods<sup>71</sup>, particle methods have several advantages. Firstly, they are easily adaptable upon modifying the model, which allows for greater flexibility in model design (in our case, useful when passing from the homogeneous description (6) to the full model with growth and cell-state transition (10), and then to the equation with generalised growth term (12)). Moreover, they are based on a Lagrangian description of the system, and do not require an

underlying mesh. More precisely, the initial data is discretised by a set of points, whose positions in the state space are then made to evolve in time via the advection term: this allows to 'follow' the cell population as it converges towards regions of higher concentration. However, in the presence of state transition terms, particle methods are not asymptotic-preserving schemes<sup>39</sup>, which means that the particle approximation (15) does not correctly approach the solution of the PDE for very large times. To avoid this problem, we carry out the regulation process at each time step, as described in the 'Methods' Section.

Overall, we employed cell population balance modelling to analyse the combined effect of EMT regulatory dynamics with cell division and death, and stochastic cell state transition. The integration of these complex processes together was made possible thanks to an efficient PDE numerical integration scheme recently developed by some of us<sup>39</sup>.

## Methods

### Parameters for ODE (8) and (9)

We detail the parameters underlying ODE (8):

$$\begin{cases} \dot{\mu}_{200} = g_{\mu_{200}} H_{Z, \mu_{200}}(Z) H_{S, \mu_{200}}(S) - g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) Q(\mu_{200}) - k_{\mu_{200}} \mu_{200} \\ \dot{Z} = g_Z g_{m_z} H_{Z, m_z}(Z) H_{S, m_z}(S) P(\mu_{200}) - k_Z Z \end{cases}$$

Functions  $H_{Z, \mu_{200}}$ ,  $H_{S, \mu_{200}}$ ,  $H_{Z, m_z}$  and  $H_{S, m_z}$  are shifted Hill functions which write under the form

$$H(X) = \frac{1 + \lambda \left( \frac{X}{X_0} \right)^n}{1 + \left( \frac{X}{X_0} \right)^n}.$$

The associated parameters are given in Table 1.

The functions  $Y_\mu$ ,  $Y_m$  and  $L$  are defined by

$$\begin{aligned} Y_\mu(\mu) &:= \sum_{i=1}^n i y_{\mu_i} \binom{n}{i} M_n^i(\mu), \quad Y_m(\mu) : \\ &= \sum_{i=1}^n y_{m_i} \binom{n}{i} M_n^i(\mu), \quad L(\mu) := \sum_{i=0}^n l_i \binom{n}{i} M_n^i(\mu), \end{aligned}$$

where  $M_n^i := \frac{(\mu/\mu_0)^i}{(1+\mu/\mu_0)^n}$ . Finally functions  $P$  and  $Q$  are given by

$$P(\mu) := \frac{L(\mu)}{Y_m(\mu) + k_{m_z}}, \quad Q(\mu) := \frac{Y_\mu(\mu)}{Y_m(\mu) + k_{m_z}}$$

Parameters for these three functions are given in Table 2.

The additional parameters of ODE (9) are  $\alpha = 0.15$ , and  $\beta(t) = 240$  when  $S$  is non-decreasing, and  $\beta(t) = 720$  when  $S$  is decreasing.

### Reduction of the advection function

Let us introduce the segments,  $I_{ep} = [150000, 185270.541082]$ ,  $I_{ep-mes} = [185270.541082, 193286.5731462]$ ,  $I_{ep-hyb-mes} = [193286.573146, 208817.635271]$ ,  $I_{hyb-mes} = [208817.635271, 224649.298597]$ , and  $I_{mes} = [224649.298597, 250000]$ , which correspond respectively to the values

**Table 1 | Parameters for the Hill functions**

				Molecules		Molecules.Hour <sup>-1</sup>		Hour <sup>-1</sup>
$n$	6			$\mu_0$	10 K			
$n_{Z, \mu_{200}}$	3	$\lambda_{Z, \mu_{200}}$	0.1	$Z_{\mu_{200}}^0$	220 K	$g_{\mu_{200}}$	2.1 K	$k_{\mu_{200}}$ 0.05
$n_{S, \mu_{200}}$	2	$\lambda_{S, \mu_{200}}$	0.1	$S_{\mu_{200}}^0$	180 K	$g_Z$	0.1 K	$k_Z$ 0.1
$n_{Z, m_z}$	2	$\lambda_{Z, m_z}$	7.5	$Z_{m_z}^0$	25 K	$g_{m_z}$	11	$k_{m_z}$ 0.5
$n_{S, m_z}$	2	$\lambda_{S, m_z}$	10	$S_{m_z}^0$	180 K			



of  $S$  for which the model is monostable (with a unique equilibrium point which corresponds to the epithelial phenotype), bistable (with two equilibrium points which correspond to the epithelial and the mesenchymal phenotypes), tristable, bistable (with two equilibrium points which correspond to the hybrid and the mesenchymal phenotypes), and monostable (with a unique equilibrium point which corresponds to the mesenchymal phenotype), as can be seen on the bifurcation diagram (Fig. 2A).

The values of stable equilibrium points representing the three possible phenotypes slightly vary depending on the value of  $S$  as illustrated by the bifurcation diagram. We approximate each of them via a five-order polynomial interpolation, i.e. a polynomial of the form

$$a_5 S^5 + a_4 S^4 + a_3 S^3 + a_2 S^2 + a_1 S + a_0.$$

$P_{mes}(S)$ ,  $P_{hyb}(S)$  and  $P_{ep}(S)$  respectively approximate the mesenchymal, hybrid and epithelial phenotypes (three solid lines on the bifurcations diagrams), while  $P_{u1}(S)$  and  $P_{u2}(S)$  correspond to the unstable equilibrium points (two dotted lines in the bifurcation diagram). The coefficients defining these five polynomials are given in Table 3.

We can finally define our one-dimensional reduced function:

- If  $S \in I_{ep}$ :  $\tilde{f}_r(x, S) = -(x - P_{ep}(S))$ .
- If  $S \in I_{ep-mes}$ :
  - If  $x \leq 0.5(P_{mes}(S) + P_{u1}(S))$ :  $\tilde{f}_r(x, S) = -(x - P_{mes}(S))$
  - If  $x \in [0.5(P_{mes}(S) + P_{u1}(S)), 0.5(P_{u1}(S) + P_{ep}(S))]$ :  $\tilde{f}_r(x, S) = x - P_{u1}(S)$
  - If  $x \geq 0.5(P_{u1}(S) + P_{ep}(S))$ :  $\tilde{f}_r(x, S) = -(x - P_{ep}(S))$
- If  $S \in I_{ep-hyb-mes}$ :
  - If  $x \leq 0.5(P_{mes}(S) + P_{u1}(S))$ :  $\tilde{f}_r(x, S) = -(x - P_{mes}(S))$
  - If  $x \in [0.5(P_{mes}(S) + P_{u1}(S)), 0.5(P_{u1}(S) + P_{hyb}(S))]$ :  $\tilde{f}_r(x, S) = (x - P_{u1}(S))$
  - If  $x \in [0.5(P_{u1}(S) + P_{hyb}(S)), 0.5(P_{hyb}(S) + P_{u2}(S))]$ :  $\tilde{f}_r(x, S) = -(x - P_{hyb}(S))$
  - If  $x \in [0.5(P_{hyb}(S) + P_{u2}(S)), 0.5(P_{u2}(S) + P_{ep}(S))]$ :  $\tilde{f}_r(x, S) = x - P_{u2}(S)$
  - If  $x \geq 0.5(P_{u2}(S) + P_{ep}(S))$ :  $\tilde{f}_r(x, S) = -(x - P_{ep}(S))$
- If  $S \in I_{hyb-mes}$ :
  - If  $x \leq 0.5(P_{mes}(S) + P_{u1}(S))$ :  $\tilde{f}_r(x, S) = -(x - P_{mes}(S))$
  - If  $x \in [0.5(P_{mes}(S) + P_{u1}(S)), 0.5(P_{u1}(S) + P_{hyb}(S))]$ :  $\tilde{f}_r(x, S) = x - P_{u1}(S)$
  - If  $x \geq \tilde{f}_r(x, S) = -(x - P_{hyb}(S))$
- If  $S \in I_{mes}$ :  $\tilde{f}_r(x, S) = -(x - P_{mes}(S))$

We are looking for a function that can be written as a multiple of  $\tilde{f}_r$ , i.e.  $f_r = k\tilde{f}_r$ , with  $k > 0$ . In order to choose the most suitable parameter, we

**Table 2 | Parameters for the functions  $L$ ,  $Y_\mu$  and  $Y_m$**

i	0	1	2	3	4	5	6
$l_i$	1.0	0.6	0.3	0.1	0.05	0.05	0.05
$\gamma_{mi}$		0.04	0.2	1.0	1.0	1.0	1.0
$\gamma_{\mu i}$		0.005	0.05	0.5	0.5	0.5	0.5

**Table 3 | Coefficients of the polynomials  $P_{mes}$ ,  $P_{u1}$ ,  $P_{hyb}$ ,  $P_{u2}$  and  $P_{ep}$**

	$P_{mes}$	$P_{u1}$	$P_{hyb}$	$P_{u2}$	$P_{ep}$
$a_5$	-6.109064e - 21	2.504295e - 19	-9.979604e - 19	1.981710e - 17	-1.980683e - 20
$a_4$	6.846339e - 15	-2.551948e - 13	1.044470e - 12	-1.996178e - 11	1.727787e - 14
$a_3$	-3.065919e - 09	1.039834e - 07	-4.372031e - 07	8.042697e - 06	-6.016556e - 09
$a_2$	6.859093e - 04	-2.117715e - 02	9.149320e - 02	-1.620159e + 00	1.045582e - 03
$a_1$	-7.668477e + 01	2.155712e + 03	-9.572462e + 03	1.631806e + 05	-9.081813e + 01
$a_0$	3.430402e + 06	-8.774602e + 07	4.005964e + 08	-6.573913e + 09	3.1850617e + 06

look for the value of  $k$  that minimises the quantity

$$\int_0^T \int_{150K}^{250K} \int_0^{25K} \int_0^{800K} |x(t, S, x_0) - \mu(t, S, x_0, Z_0)| dt dS dx_0 dZ_0, \quad (13)$$

where for all  $x_0 \in [0, 25K]$ ,  $S \in [150K, 250K]$ ,  $x(\cdot, S, x_0)$  solves

$$\begin{cases} \dot{x}(t, S, x_0) = f_r(x(t, S, x_0)) \\ x(0, S, x_0) = x_0 \end{cases},$$

and for all  $Z_0 \in [0, 800K]$ ,  $(\mu(\cdot, S, x_0, Z_0), Z(\cdot, S, x_0, Z_0))$  solves (8).

In practice, this integral has been approximated for  $T \in \{10, 100, 1000\}$  by the Riemann sum

$$\frac{1}{N_T N_S N_x N_Z} \sum_{i,j,k,l} |x(t_i, S_j, x_{0k}) - \mu(t_i, S_j, x_{0k}, Z_{0l})|,$$

with  $N_S = N_x = N_Z = 20$ ,  $N_T = 100$ , and for all  $i, j, k, l$ ,  $t_i = i \frac{T}{N_T}$ ,  $S_j = 150K + j \frac{100K}{N_S}$ ,  $x_{0k} = k \frac{25K}{N_x}$ , and  $Z_{0l} = l \frac{800K}{N_Z}$ , where  $x(\cdot, S_j, x_{0k})$  solves

$$\begin{cases} \dot{x}(t, S_j, x_{0k}) = f_r(x(t, S_j, x_{0k})) \\ x(0, S_j, x_{0k}) = x_{0k} \end{cases},$$

$(\mu(\cdot, S_j, x_{0k}, Z_{0l}), Z(\cdot, S_j, x_{0k}, Z_{0l}))$  solves (8).

The value of the multiplicative constant which minimises (13) depends on  $T$  but we establish that, for  $T \in \{10, 100, 1000\}$  its value is rather insensitive to that of  $T$ : it is close to 0.02, which is the value that we select. The obtained function is shown in Fig. S2 for various values of  $S$ .

## Numerical method

We start by normalising the model to work with the domain  $[0, 1] \times [0, 1]$  rather than  $[0, 25K] \times [150K, 250K]$ . By denoting  $A = 25K$ ,  $B = 0$ ,  $C = 100K$ ,  $D = 150K$ , we check that for all  $t \geq 0$ ,  $u(t, x, S) = \frac{1}{AC} u_{re}(\frac{x-B}{A}, \frac{S-D}{C})$ , where  $u_{re}$  is the solution of

$$\begin{cases} \partial_t u_{re}(t, y) + \nabla \cdot (f_{re}(y) u_{re}(t, y)) = (r_{re}(y) - d_{re}(y) \rho(t)) u_{re}(t, y) \\ \quad + \int_{\mathbb{R}^2} M_{re}(y, z) u(t, z) dz - \int_{\mathbb{R}^2} M_{re}(z, y) dz u(t, y) \\ \rho(t) = \int_{\mathbb{R}^2} u_{re}(t, y) dy \end{cases}, \quad (14)$$

where for all  $x, S \in \mathbb{R}$ ,  $f_{re}(x, S) := (\frac{1}{A} f_r(Ax + B, CS + D), \frac{1}{C} f_s(CS + D))$ ,  $r_{re}(x, S) := r(Ax + B, CS + D)$ ,  $d_{re}(x, S) := d(Ax + B, CS + D)$ ,  $M_{re}(x, S, x', S') := ACM(Ax + B, CS + D, Ax' + B, CS' + D)$ , and  $u_{re}^0(x, S) := ACu^0(Ax + B, CS + D)$ .

Thus, an approximation for  $u_{re}$  provides an approximation for  $u$ . We make use of a particle method in order to approximate  $u_{re}$  at different times ( $0 < T_1 < \dots < T_K$ , specified in each figure), applying the method introduced in ref. 39 to deal with a category of models to which (10) belongs. For a proof that the numerical scheme does successfully approximate the solutions of

(10), we refer to ref. 39, while an introduction to particle methods can be found in ref. 38.

We choose an integer parameter  $N$  ( $N = 20$  in our simulations), and we denote  $y_1^0, \dots, y_{N^2}^0$ , the points of the grid of size  $N \times N$  on  $[0, 1] \times [0, 1]$ . For  $i \in \{1, \dots, N^2\}$ , we solve the ODE

$$\begin{cases} \dot{y}_i = f(y_i) \\ \dot{w}_i = \nabla \cdot f(y_i) w_i \\ \dot{v}_i = \left( r(y_i) - d(y_i) \sum_{j=1}^{N^2} v_j \right) v_i + w_i \sum_{j=1}^{N^2} M(y_i, y_j) v_j - \sum_{j=1}^{N^2} M(y_j, y_i) v_i \end{cases}, \quad (15)$$

with initial conditions  $y_i(0) = y_i^0$ ,  $w_i(0) = \frac{1}{N^2}$  and  $v_i(0) = \frac{u_\varepsilon^0(y_i^0)}{N^2}$ , on the first time interval  $([0, T_1])$ . The solution of this ODE is called *Particle approximation of (10)*. To solve (15) on  $[0, T_1]$ , we use the Python function `solve_ivp` in module `scipy.integrate`, with the default solver which corresponds to an explicit Runge-Kutta method of order 5<sup>72</sup>.

We then use a *regularisation process*, i.e. we compute the sum

$$u_\varepsilon^N(T_1, x) = \sum_{i=1}^{N^2} v_i(T_1) G_\varepsilon(y - y_i(T_1)), \quad (16)$$

with  $G_\varepsilon(x, S) := \frac{1}{\varepsilon^2} G(\frac{x}{\varepsilon}) G(\frac{S}{\varepsilon})$ , with  $G$  the Gaussian function, and  $\varepsilon = (\frac{1}{N^2})^\gamma$ , with  $\gamma \in (0.5, 1)$ . In all our simulations, we choose  $\gamma = 0.8$ : this value has been chosen empirically by carrying out simulations in simple cases for which the behaviour of solutions is well known (for example in the absence of cell-state transition), and by comparing with other values of  $\gamma$ . The points ‘ $y$ ’ at which we compute this sum are  $y_1^0, \dots, y_{N^2}^0$ .

We repeat the process for each time interval, taking as initial data the approximation calculated at the previous time step (i.e.  $u^N(T_{k-1}, \cdot)$ ), to compute  $u^N(T_k, \cdot)$ .

We acknowledge that stochastic approaches might also be used to simulate such models.

## Data availability

The codes used in manuscript to generate all simulation data are available here: [https://github.com/Paras-Jain20/EMT\\_Cell\\_Population\\_Balance](https://github.com/Paras-Jain20/EMT_Cell_Population_Balance).

Received: 25 June 2024; Accepted: 17 February 2025;

Published online: 06 March 2025

## References

- Jacquemin, V. et al. Dynamic cancer cell heterogeneity: diagnostic and therapeutic implications. *Cancers* **14**, 280 (2022).
- Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer Cell* **37**, 471–484 (2020).
- Cross, W. et al. The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* **2**, 1661–1672 (2018).
- Andor, N. et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
- Nowell, P. C. The clonal evolution of tumor cell populations: acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. *Science* **194**, 23–28 (1976).
- Bell, C. & Gilan, O. Principles and mechanisms of non-genetic resistance in cancer. *Br. J. Cancer* **122**, 465–472 (2020).
- Pillai, M., Hojel, E., Jolly, M. & Goyal, Y. Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. *Nat. Comput. Sci.* **2023** 3:4 **3**, 301–313 (2023).
- Goyal, Y. et al. Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. *Nature* **620**, 651–659 (2023).
- Marusyk, A. et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58 (2014).
- Karacosta, L. et al. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nat. Commun.* **10**, 5587 (2019).
- Cook, D. P. & Vanderhyden, B. C. Context specificity of the emt transcriptional response. *Nat. Commun.* **11**, 2142 (2020).
- Ruscetti, M. et al. Hdac inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. *Oncogene* **35**, 3781–3795 (2016).
- Subbalakshmi, A. R. et al. Nfatc acts as a non-canonical phenotypic stability factor for a hybrid epithelial/mesenchymal phenotype. *Front. Oncol.* **10**, 553342 (2020).
- Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).
- Risom, T. et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat. Commun.* **9**, 1–17 (2018).
- Font-Clos, F., Zapperi, S. & Porta, C. Topography of epithelial-mesenchymal plasticity. *Proc. Natl Acad. Sci. USA* **115**, 5902–5907 (2018).
- Hari, K. et al. Identifying inhibitors of epithelial-mesenchymal plasticity using a network topology-based approach. *Npj Syst. Biol. Appl.* **6**, 1–12 (2020).
- Hong, T. et al. An oval2-zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* **11**, 1004569 (2015).
- Rashid, M., Hari, K., Thampi, J., Santhosh, N. & Jolly, M. Network topology metrics explaining enrichment of hybrid epithelial/mesenchymal phenotypes in metastasis. *PLoS Comput. Biol.* **18**, 1010687 (2022).
- Steinway, S. et al. Combinatorial interventions inhibit tgfb-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *Npj Syst. Biol. Appl.* **1**, 15014 (2015).
- George, J., Jolly, M., Xu, S., Somarelli, J. & Levine, H. Survival outcomes in cancer patients predicted by a partial emt gene expression scoring metric. *Cancer Res.* **77**, 6415–6428 (2017).
- Brown, M. et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Sci. Adv.* **8**, eabj8002 (2022).
- Celià-Terrassa, T. et al. Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat. Commun.* **9**, 5005 (2018).
- Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A. & Jolly, M. Landscape of epithelial-mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. *ELife* **11**, e76535 (2022).
- Bintu, L. Dynamics of epigenetic regulation at the single-cell level. *Biophys. J.* **110**, 317a–318a (2016).
- Hathaway, N. A. et al. Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012).
- Boareto, M. et al. Notch-jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J. R. Soc. Interface* **13**, 20151106 (2016).
- Jolly, M. et al. Inflammatory breast cancer: A model for investigating cluster-based dissemination. *Npj Breast Cancer* **3**, 1–7 (2017).
- Neelakantan, D. et al. Emt cells increase breast cancer metastasis via paracrine gli activation in neighbouring tumour cells. *Nat. Commun.* **8**, 15773 (2017).
- Yamamoto, M. et al. Intratumoral bidirectional transitions between epithelial and mesenchymal cells in triple-negative breast cancer. *Cancer Sci.* **108**, 1210–1222 (2017).
- Hitomi, M. et al. Asymmetric cell division promotes therapeutic resistance in glioblastoma stem cells. *JCI Insight* **6**, e130510 (2021).
- Tripathi, S., Chakraborty, P., Levine, H. & Jolly, M. A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. *PLoS Comput. Biol.* **16**, 1–27 (2020).

33. Munsky, B., Fox, Z. & Neuert, G. Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. *Methods* **85**, 12–21 (2015).
34. Pally, D., Goutham, S. & Bhat, R. Extracellular matrix as a driver for intratumoral heterogeneity. *Phys. Biol.* **19**, 043001 (2022).
35. Lovisa, S. et al. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nat. Med.* **21**, 998–1009 (2015).
36. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in trail-induced apoptosis. *Nature* **459**, 428–432 (2009).
37. Strasen, J. et al. Cell-specific responses to the cytokine  $\text{tgf } \beta$  are determined by variability in protein levels. *Mol. Syst. Biol.* **14**, e7733 (2018).
38. Chertock, A. A practical guide to deterministic particle methods. In *Handbook of Numerical Analysis* Vol. 18, 177–202 (Elsevier, 2017).
39. Alvarez, F. E. & Guilberteau, J. A particle method for non-local advection–selection–mutation equations. *Math. Models Methods Appl. Sci.* **34**, 597–658 (2024).
40. Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proc. Natl Acad. Sci. USA* **110**, 18144–18149 (2013).
41. Jain, P., Bhatia, S., Thompson, E. & Jolly, M. Population dynamics of epithelial–mesenchymal heterogeneity in cancer cells. *Biomolecules* **12**, 348 (2022).
42. Jia, W., Deshmukh, A., Mani, S. A., Jolly, M. K. & Levine, H. A possible role for epigenetic feedback regulation in the dynamics of the epithelial–mesenchymal transition (EMT). *Phys. Biol.* **16**, 066004 (2019).
43. Jain, P. et al. Epigenetic memory acquired during long-term emt induction governs the recovery to the epithelial state. *J. R. Soc. Interface* **20**, 20220627 (2023).
44. Miyamoto, T., Furusawa, C. & Kaneko, K. Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. *PLoS Comput. Biol.* **11**, e1004476 (2015).
45. Bhatia, S. et al. Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. *J. Clin. Med.* **8**, 893 (2019).
46. Sigal, A. et al. Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
47. Vega, S. et al. Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **18**, 1131–1143 (2004).
48. Subbalakshmi, A. R., Sahoo, S., Biswas, K. & Jolly, M. K. A computational systems biology approach identifies slug as a mediator of partial epithelial–mesenchymal transition (emt). *Cells Tissues Organs* **211**, 689–702 (2022).
49. Subbalakshmi, A. R. et al. The *elf3* transcription factor is associated with an epithelial phenotype and represses epithelial–mesenchymal transition. *J. Biol. Eng.* **17**, 17 (2023).
50. Silveira, D. A., Gupta, S. & Mombach, J. C. M. Systems biology approach suggests new mirnas as phenotypic stability factors in the epithelial–mesenchymal transition. *J. R. Soc. Interface* **17**, 20200693 (2020).
51. Kröger, C. et al. Acquisition of a hybrid e/m state is essential for tumorigenicity of basal breast cancer cells. *Proc. Natl Acad. Sci. USA* **116**, 7353–7362 (2019).
52. Corre, G. et al. Stochastic fluctuations and distributed control of gene expression impact cellular memory. *PLoS One* **9**, e115574 (2014).
53. Westbrook, E. R., Ford, H. Z., Antolović, V. & Chubb, J. R. Clearing the slate: RNA turnover to enable cell-state switching? *Development* **150**, dev202084 (2023).
54. Stylianou, N. et al. A molecular portrait of epithelial–mesenchymal plasticity in prostate cancer associated with clinical outcome. *Oncogene* **38**, 913–934 (2019).
55. Farrokhi, N. et al. Measuring competitive exclusion in non-small cell lung cancer. *Sci. Adv.* **8**, eabm7212 (2022).
56. West, J. B. et al. Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: an evolution-based strategy. *Clin. Cancer Res.* **25**, 4413–4421 (2019).
57. Noble, R. J. et al. Paracrine behaviors arbitrate parasite-like interactions between tumor subclones. *Front. Ecol. Evol.* **9**, 675638 (2021).
58. Deng, Y., Chakraborty, P., Jolly, M. K. & Levine, H. A theoretical approach to coupling the epithelial–mesenchymal transition (EMT) to extracellular matrix (ECM) stiffness via *loxl2*. *Cancers* **13**, 1609 (2021).
59. Mantzaris, N. From single-cell genetic architecture to cell population dynamics: Quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. *Biophys. J.* **92**, 4271–4288 (2007).
60. Hasenauer, J. et al. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinform.* **12**, 1–15 (2011).
61. Shu, C., Chatterjee, A., Dunny, G., Hu, W. & Ramkrishna, D. Bistability versus bimodal distributions in gene regulatory processes from population balance. *PLoS Comput. Biol.* **7**, e1002140 (2011).
62. Spetsieris, K., Zygorakis, K. & Mantzaris, N. A novel assay based on fluorescence microscopy and image processing for determining phenotypic distributions of rod-shaped bacteria. *Biotechnol. Bioeng.* **102**, 598–615 (2009).
63. Hasenauer, J., Schittler, D. & Allgöwer, F. Analysis and simulation of division-and label-structured population models: a new tool to analyze proliferation assays. *Bull. Math. Biol.* **74**, 2692–2732 (2012).
64. Schittler, D., Allgöwer, F. & De Boer, R. J. A new model to simulate and analyze proliferating cell populations in BrdU labeling experiments. *BMC Syst. Biol.* **7**, 1–6 (2013).
65. Loos, C., Moeller, K., Fröhlich, F., Hucho, T. & Hasenauer, J. A hierarchical, data-driven approach to modeling single-cell populations predicts latent causes of cell-to-cell variability. *Cell Syst.* **6**, 593–603 (2018).
66. Mantzaris, N. V. Stochastic and deterministic simulations of heterogeneous cell population dynamics. *J. Theor. Biol.* **241**, 690–706 (2006).
67. Degond, P. & Mas-Gallic, S. The weighted particle method for convection–diffusion equations. i. the case of an isotropic viscosity. *Math. Comput.* **53**, 485–507 (1989).
68. Waldherr, S. Estimation methods for heterogeneous cell population models in systems biology. *J. R. Soc. Interface* **15**, 20180530 (2018).
69. Spetsieris, K. & Zygorakis, K. Single-cell behavior and population heterogeneity: solving an inverse problem to compute the intrinsic physiological state functions. *J. Biotechnol.* **158**, 80–90 (2012).
70. Mantzaris, N. V., Daoutidis, P. & Sreenc, F. Numerical solution of multi-variable cell population balance models. III. finite element methods. *Comput. Chem. Eng.* **25**, 1463–1481 (2001).
71. Mantzaris, N. V., Daoutidis, P. & Sreenc, F. Numerical solution of multi-variable cell population balance models: I. finite difference methods. *Comput. Chem. Eng.* **25**, 1411–1440 (2001).
72. Dormand, J. R. & Prince, P. J. A family of embedded runge–kutta formulae. *J. Comput. Appl. Math.* **6**, 19–26 (1980).

## Acknowledgements

This work was supported by the Raman–Charpak Fellowship 2022 awarded to J.G., funding a 2-month research stay at the Indian Institute of Science, Bangalore. M.K.J. was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) awarded by Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India. N.P.D. was supported by Emergence fellowship (S21JR31024) awarded by Sorbonne University.



## Author contributions

J.G. and P.J. performed research. M.K.J., C.P. and N.P.D. designed and supervised research. All authors contributed to data analysis and in writing and reviewing the manuscript.

## Competing interests

M.K.J. serves as Editor-in-Chief of this journal and had no role in the peer-review or decision to publish this manuscript. M.K.J. declares no financial competing interests. All other authors declare no financial or non-financial competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41540-025-00502-4>.

**Correspondence** and requests for materials should be addressed to Mohit Kumar Jolly, Camille Pouchol or Nastassia Pouradier Duteil.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025