HAL
open science

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

Jules Guilberteau, Paras Jain, Mohit Kumar Jolly, Camille Pouchol, Nastassia Pouradier Duteil

To cite this version:

Jules Guilberteau, Paras Jain, Mohit Kumar Jolly, Camille Pouchol, Nastassia Pouradier Duteil. An integrative phenotype-structured partial differential equation model for the population dynamics of epithelial-mesenchymal transition. 2023. hal-04208893

HAL Id: hal-04208893

https://hal.science/hal-04208893

Preprint submitted on 15 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Jules Guilberteau[1,4], Paras Jain[2,4], Mohit Kumar Jolly[2,5], Camille Pouchol[3,5], and Nastassia Pouradier Duteil[1,5]

[1]Sorbonne Université, CNRS, Université Paris Cité, Inria, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France.
[2]Department of Bioengineering, Indian Institute of Science, Bangalore 560012, India
[3]Université Paris Cité, FP2M, CNRS FR 2036, MAP5 UMR 8145, F-75006 Paris, France.

[4]Equally contributed to the work
[5]To whom correspondence may be addressed: M.K.J.: mkjolly@iisc.ac.in, C.P.: camille.pouchol@u-paris.fr, N.P.D.: nastassia.pouradier_duteil@sorbonne-universite.fr

September 15, 2023

Abstract

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 possible processes interplay in shaping the dynamics of E-M heterogeneity: a) intracellular regulatory interaction among biomolecules, b) cell division and death, and c) stochastic cell-state transition (biochemical reaction noise and asymmetric cell division). Here, we propose a Cell Population Balance (Partial Differential Equation (PDE)) based model that captures the dynamics of cell population density along the E-M phenotypic axis due to abovementioned multi-scale cellular processes. We demonstrate how population distribution resulting from intracellular regulatory networks driving cell-state transition gets impacted by stochastic fluctuations in E-M regulatory biomolecules, differences in growth rates among cell subpopulations, and initial population distribution. Further, we reveal that a linear dependence of the cell growth rate on the population heterogeneity is sufficient to recapitulate the faster *in vivo* growth of orthotopic injected heterogeneous E-M subclones reported before experimentally. Overall, our model contributes to the combined understanding of intracellular and cell-population levels dynamics in the emergence of E-M heterogeneity in a cell population.

1 Introduction

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 [1, 2]. Such heterogeneity can arise at a genetic level over the course of tumour evolution and manifests as different clonal populations. However, over the last two decades, non-genetic or phenotypic heterogeneity among cancer cells has been identified as a key driver of disease aggressiveness [3, 4]. 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 spectrum. Given the implications of E-M heterogeneity

in cancer metastasis and patient outcomes, various in vitro, in vivo and in silico attempts have been focused on understanding its underlying mechanisms [5, 6, 7, 8].

An iterative cross-talk among *in silico* and *in vitro* studies has contributed enormously to understanding 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) cell states [9, 10, 11, 12, 13]. Their co-existence has been experimentally reported in varying ratios in multiple cell lines and primary tumours [14, 7]. 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 [15, 5, 16, 17]. The relative stability of cells in different phenotypes, and consequently the phenotypic distribution at a given time, is governed by the underlying topology of the regulatory network involving transcriptional and translational control [18], as well as by epigenetic regulation (chromatin modification). 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 phenotypic distribution. Further, cell-to-cell communications either through neighbourhood interactions or paracrine signalling also reshapes the phenotypic distribution [19, 20, 21, 22]. Hence, a diversity of regulatory interactions within and among cells can contribute to shaping the E-M heterogeneity patterns in a cell population.

Another milieu of factors such as asymmetric cell division [23, 24], stochastic biochemical noise [25], differences in cellular microenvironment [26], and variable cell-cycle dynamics [27] 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 [28, 29] 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. Broadly speaking, two modelling approaches are employed:

1. Agent-Based Cell Population models: population models that capture regulatory and stochastic dynamics for individual cells and then generate an ensemble to generate population distribution [6, 30, 24].

2. Cell Population Balance models: population models that capture the regulatory dynamics and stochastic dynamics of groups of cells with similar cell-states without dealing with individual cell level information.

The latter approach has been adopted widely because its output (cell density) can be related directly to flow cytometry experiments conducted at multiple timepoints for a population. These models describe the evolution of the cell density in the phenotypic space. More precisely, they track the number of cells that have a given phenotype in the E-M landscape, where the phenotype is defined by a vector containing the concentration of relevant molecules determining the phenotype of a cell. Further, Cell Population Balance models have been used to combine complex regulatory phenomena, e.g. positive feedback loops [31], caspase activation cascade during programmed cell death [32], cell-to-cell communication [33], and two mutually inhibiting nodes [34], with stochastic processes like asymmetric cell division and stochastic biochemical reactions for phenotype-structured, and age-structured cell populations [35, 36, 33]. Finally, since Agent-Based Cell Population model simulate dynamics of each cell individually, they become computationally intractable for realistically large numbers of cells, contrary to models for cell densities.

Therefore, given its wide applicability, we have here used Cell Population Balance modelling to study the population dynamics of E-M heterogeneity. 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 aforementioned molecules, 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 demonstrate how heterogeneity along the Epithelial-Mesenchymal axis emerges at a population level. First, we 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). Second, we report how cellular heterogeneity depends on the characteristic of stochastic biochemical noise in – 1)

external EMT inducing signal (SNAIL), and 2) E-M state variable (miR200 or ZEB levels) – along with the differences in the relative growth rate among E, E/M and M states, and the initial distribution of population in the EMT cell state spectrum. Finally, we show that a population with its fitness (growth rate) proportional to the heterogeneity of population (Renyi entropy) can explain faster tumour growth in vivo and higher proliferation rates in vitro as observed for parental and intermediate clones derived from SUM149T and PMC42-LA cell lines, respectively [37, 5].

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) [38]. 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 [39], where it is proved that solutions are properly approximated by the proposed numerical method.

2 Results

2.1 Building a Cell Population Balance Model

Cell Population Balance models allow to compute the evolution of the number of cells of phenotype $y \in \mathbb{R}^n$ at time t, denoted by $u(t, y)$. Here, n represents the dimension of the phenotypic space, that we will also refer to as "state space". In our context, a cell's phenotype is considered to be determined by the concentration of several different (internal or external) molecules: the dimension n of the phenotypic space will then correspond to the number of different molecules that determine its phenotype.

To give an intuition for the final PDE model, we consider the simplest 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 phenotypes belong 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, but for synthetic purposes, from here onward, we will refer to u as the *cell density*.

In practice, as will be seen in the subsequent sections, two molecules, that are markers of EMT, will be taken into account for the description of the (two-dimensional) phenotypic space: miR200 and ZEB. The phenotypic space will then be reduced to one dimension by empirically introducing an artificial state variable x, roughly equivalent to miR200. Then, the cell density u will be expressed in units of cells per number of molecules. As an example, to compute the number of cells which contain fewer than 500 molecules x, we compute the integral $\int_0^{500} u(t, y) dy$. With the previous notations, $a = 0$ and $b = 500$.

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.

Growth. The first mechanism that we include in our model is **growth**, which takes into account 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 intrinsic growth 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 $t + \Delta t$ caused by the cell population growth is given by

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

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).

State Regulation. We take advantage of ODE models built to describe the evolution of the phenotype y of one given cell, which takes the form $\dot{y} = f(y)$. 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).

Stochastic cell transition. The third mechanism that we take into account is stochastic cell-state transition, that we will also refer to as **mutations**. 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 $\Delta t \int_a^b \left(\int_{\mathbb{R}} M(y, z)u(t, z)\, dz \right) dy$. 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 $\Delta t \int_a^b \left(\int_{\mathbb{R}} M(z, y)\, dz \right) u(t, y)\, dy$. 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).

Resulting PDE model. Putting everything together, the variation of cells whose phenotype belongs to $[a, b]$ during a time interval Δt is approximated by

$$\begin{aligned}\int_a^b u(t + \Delta t, y)\, dy - \int_a^b u(t, y)\, dy = \Delta t \Bigg[& \underbrace{\int_a^b (r(y) - d(y)\rho(t))\, u(t, y)\, dy}_{\text{growth}} \\ & + \underbrace{f(a)u(t, a) - f(b)u(t, b)}_{\text{advection}} \\ & + \underbrace{\int_a^b \left(\int_{\mathbb{R}} M(y, z)u(t, z)\, dz \right) dy - \int_a^b \left(\int_{\mathbb{R}} M(z, y)\, dz \right) u(t, y)\, dy}_{\text{mutations}} \Bigg].\end{aligned}$$

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

Taking the limit Δt going to zero, the partial differential equation modelling the mechanisms of advection, growth and mutations is then written (for any dimension n) as

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

which must be complemented with an initial condition describing the density of cells at time 0.

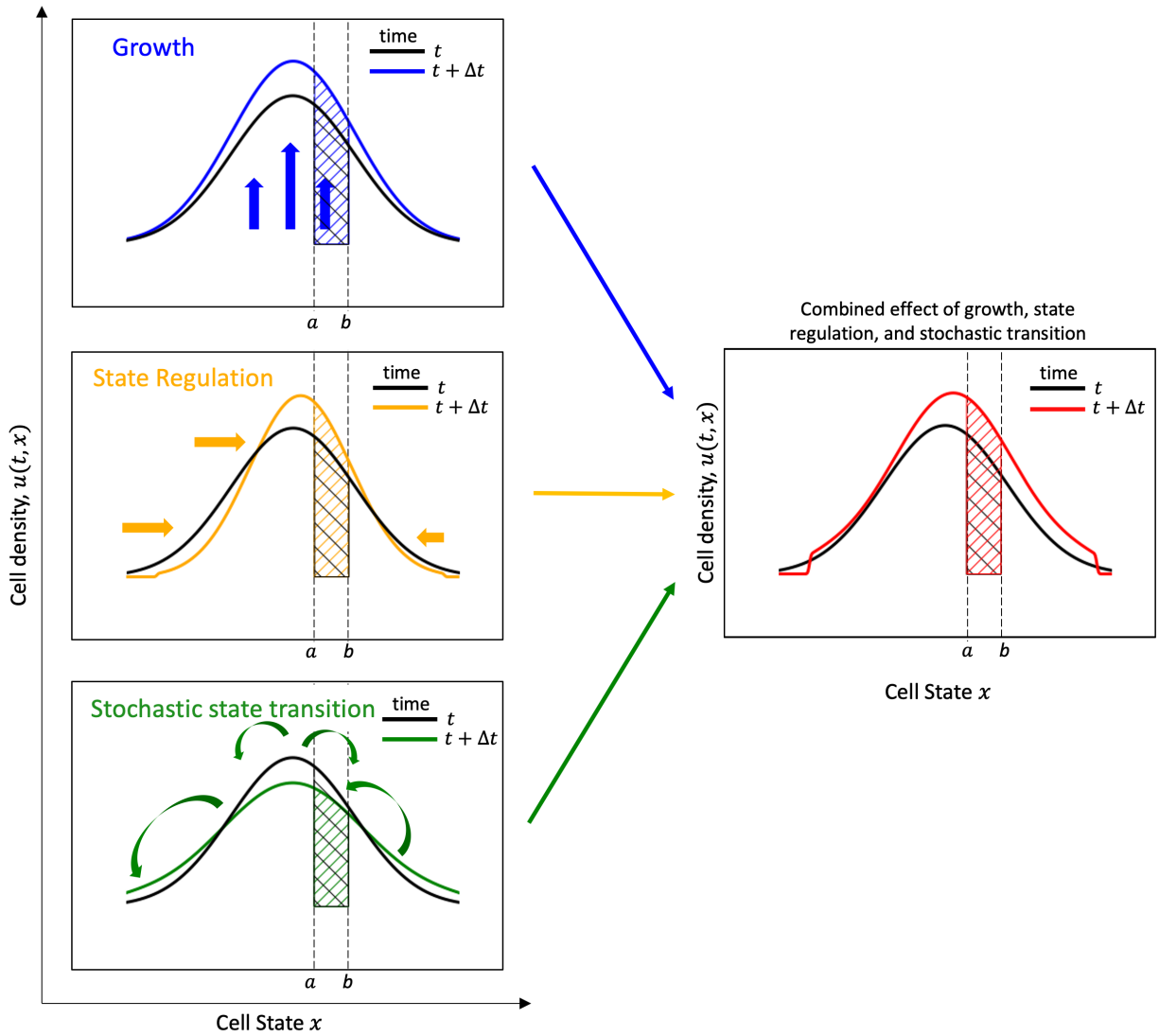

Figure 1: Growth, state regulation and stochastic state transition combine to give the cell density evolution in the proposed Cell Population Balance model. In each panel, the cell density u is represented at time t in black, and at a subsequent time $t + \Delta t$ in color. The number of cells with cell state in the interval $[a, b]$ is represented by the hashed (black or colored) area. Growth (which includes cell division and cell death) results in an upward or downward 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 cell transition is 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).

2.2 A Cell Population Balance Model recapitulates key dynamical aspects of EMT

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 hybrid E/M states.

To properly define the advection function f underlying state regulation, we chose a minimal EMT regulatory network with canonical epithelial (microRNA-200 (miR200)) 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 (Figure 2 A) [40]. The bifurcation diagram depicts the different possible stable states, each characterised by a specific range of miR200 levels (solid lines) for increasing levels of SNAIL, resulting from the network dynamics. As a cell undergoes EMT (i.e. SNAIL levels increase), it switches from high to intermediate to low levels of miR200 which corresponds to E, E/M and M state respectively. However, during MET, the cell switches directly from low (M) to high (E) miR200 levels, thus displaying hysteresis.

We confirmed that the Cell Population Balance model developed here captures hysteresis, upon neglecting cell growth and transition (biochemical noise) and using SNAIL dynamics shown in black curve (Figure 2 B, see Methods Section for formalism). We simulated the dynamics for homogenous and heterogenous cell population with respect to their distribution of SNAIL levels. For a homogeneous cell population (Figure 2 C,i), we saw that the cells reside in three distinct miR200 states (high, intermediate, and low) for time 0 to 5000 hours (increasing SNAIL levels), but made a quick transition from low to high miR200 levels for time 5000 to 10000 hours without spending much time in intermediate state (decreasing SNAIL levels). The intermediate miR200 levels seen during MET are to be understood as a sample timepoint where miR200 levels are responding to changes in SNAIL levels before settling to its equilibrium (high) state. Similar observation of hysteresis was made while considering a heterogeneous cell population (Figure 2 C, ii). Particularly, the cell distribution along ZEB and miR200 axis during MET shows that the transient intermediate miR200 peak in homogeneous population has turned into little dispersed transient peaks which were clearly distinct from the intermediate peak arising during EMT (Figure S1). This observation recapitulates the experimental data on different partial states seen during EMT vs. during MET [7]. 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 (Figure 2 C, ii).

The range of SNAIL values for which E, E/M and M states are stable (Figure 2 A) can be altered by epigenetic (chromatin-based) changes that can happen during long-term EMT induction, leading to a delayed MET [6, 41]. Therefore, to capture this phenomenon within our modelling framework, we incorporated the phenomenological formalism to account for epigenetic changes in EMT regulatory network (equation (4), [30]). Again, we neglected cell growth and transitions (biochemical noise) for this analysis. We consider two SNAIL levels dynamics to show the influence of epigenetic changes during EMT: Short-term induction (Figure 2 B blue curve), and Long-term induction (Figure 2 B red curve). We saw that homogeneous cell population have delayed recovery time for long-term EMT induction than short-term induction, therefore, recapitulating the previous observation based on population's average cell analysis [30].

Overall, the developed Cell Population Balance Model hence captures the dynamical features associated with EMT/MET.

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

Each individual cell's state (characterised by levels of a set of specific biomolecules) can dynamically evolve due to stochasticity in biochemical reactions or cell division, thus causing heterogeneity. We focus on the latter case, that is we consider that stochastic cell transitions occur at cell division.

To evaluate how these stochastic processes influence the population distribution of E, M and E/M states, we next observe cell state distribution as a result of stochastic cell-state transition in a regulatory network

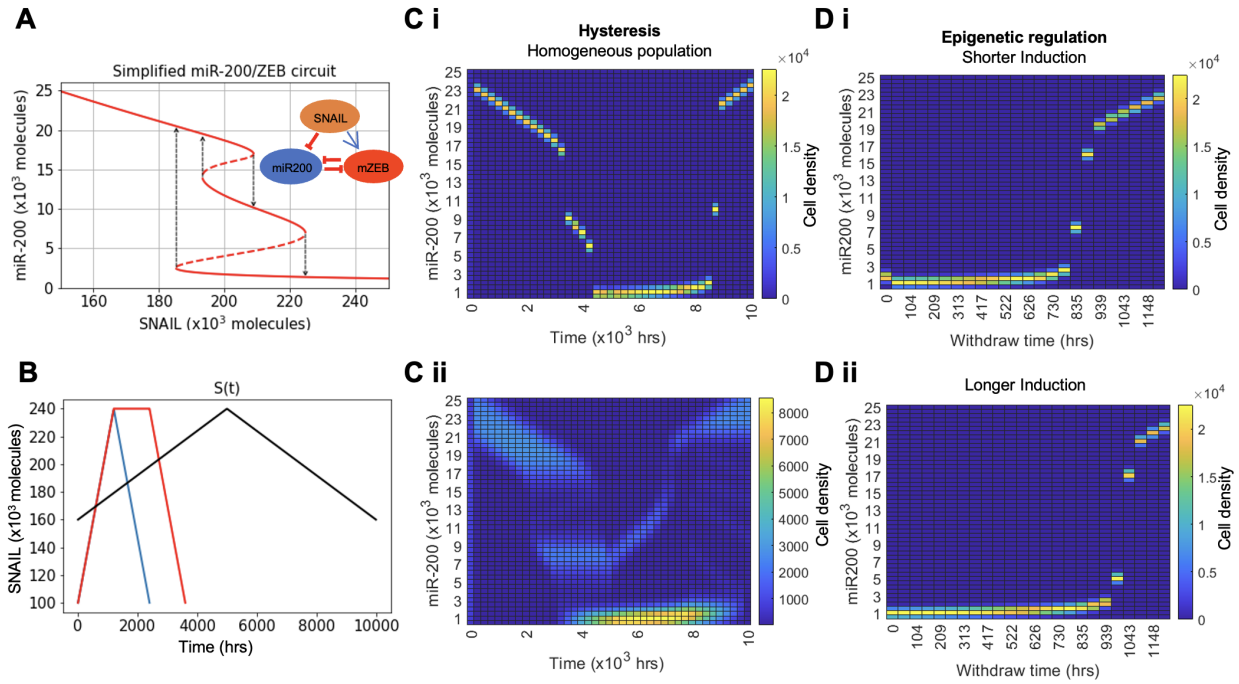

Figure 2: Proposed cell population balance model reproduces hysteresis and epigenetic regulation of state transitions at the population levels. A) EMT gene regulatory network (inset) and the bifurcation of cell states resulting from the network dynamics with increasing input signal (SNAIL) levels, B) Variation in the external input (SNAIL) levels with time to capture hysteresis and epigenetic regulation of cell states in C and 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 Figure 2B while considering – i) Homogeneous population and ii) Heterogeneous population. D) Dynamics of miR200 levels in homogeneous population during withdrawal of SNAIL levels post short and long-term EMT induction (blue and red curve in Figure 2B, respectively). Growth of cells and state transition because of biochemical noise are neglected in these simulation results.

for a population where cell division rate is independent of cell-state (i.e. assuming EMT does not impact cell cycle) (equation (5)) and is uniform for all three E, M and hybrid E/M subpopulations. 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 (Figure 3 A, i). 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 η_x and η_S, respectively.

To perform a comprehensive analysis of the impact of stochastic cell-state transition in a computationally efficient manner, we approximated the two-dimensional ODE (with variables miR200 and ZEB) in the EMT network by a one-dimensional ODE satisfied by a variable denoted x. The variable x in the reduced system is built to be roughly equivalent to levels of miR200 in the full EMT network (more details about model reduction is presented in Appendix C).

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. The Methods section and Appendix C mention the empirical formulation of the reduced system and its optimal parameterisation to minimise the error in dynamical results obtained using complete vs reduced EMT system characteristics. We first established similarity between the dynamics of the system with two variables (full EMT) and that of the reduced system by comparing the distribution of miR200 levels for a given SNAIL

distribution characteristic with the cell population distribution along the state 'x' (Figure 3 A and Figure S2 B). For example, with a SNAIL distribution of mean value (S_0) of 200K molecules, the bifurcation diagram depicts the possibility of cell population to be distributed in all three states (Figure 3 Ai). We observed that the asymptotic distribution of state variable x from reduced system dynamics exhibits tri-modality, showing the co-existence of all three states, irrespective of initial population condition (Figure 3 A, ii – initial population: all cells as epithelial, Figure S2 B – initial population: all cells as hybrid or mesenchymal). Similarly for SNAIL distribution with mean $S_0 \in \{190K, 225K, 150K, 250K\}$ molecules, we observed respective combinations of phenotypes as in bifurcation diagram of full EMT network – co-existence of E & E/M (bimodal), co-existence of hybrid E/M and M (bimodal), epithelial (unimodal) and mesenchymal (unimodal) (Fig 3 A), thus providing further evidence of faithfully representing EMT dynamics through the sole variable x.

For multi-modal distributions of 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 hybrid E/M (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 η_x of stochastic cell-state transition in state x (Figure 3 Bi, ii). Similarly, for distribution of SNAIL levels with mean S_0 = 175K molecules, although both E and M states co-exist (Figure 3 A, 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 η_x (Figure 3 B, iii).

As mentioned previously, the distribution of SNAIL levels and correspondingly that of 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 hybrid E/M prostate cancer subpopulations are segregated, they exhibit very different distributions after two weeks [15]. Similarly, the segregated EpCAM-high and EpCAM-low subpopulations in breast cancer have varied recovery dynamics [37]. Thus, in case of either internal (stochastic cell-state transition) or external (microenvironmental 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 proportional the parameter α in our model, which captures the characteristic time of convergence of SNAIL levels to its equilibrium S_0 [30, 42]. 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 α (Figure 3 C). 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 Renyi entropy (Figure 3 D). 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 turns to a majority of M cells. Similar observations can be made for other combinations of initial condition and mean S_0 values of SNAIL distribution (Figure S3, Figure 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.

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

We demonstrate 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 hybrid E/M) to divide at equal rates. To observe the additional influence of growth 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 hybrid E/M cells divide at equal rates, while M cells divide at half the rate of E cells; and c) case 'r3': both E/M and M divide at equal rates, which is half the rate of division of E

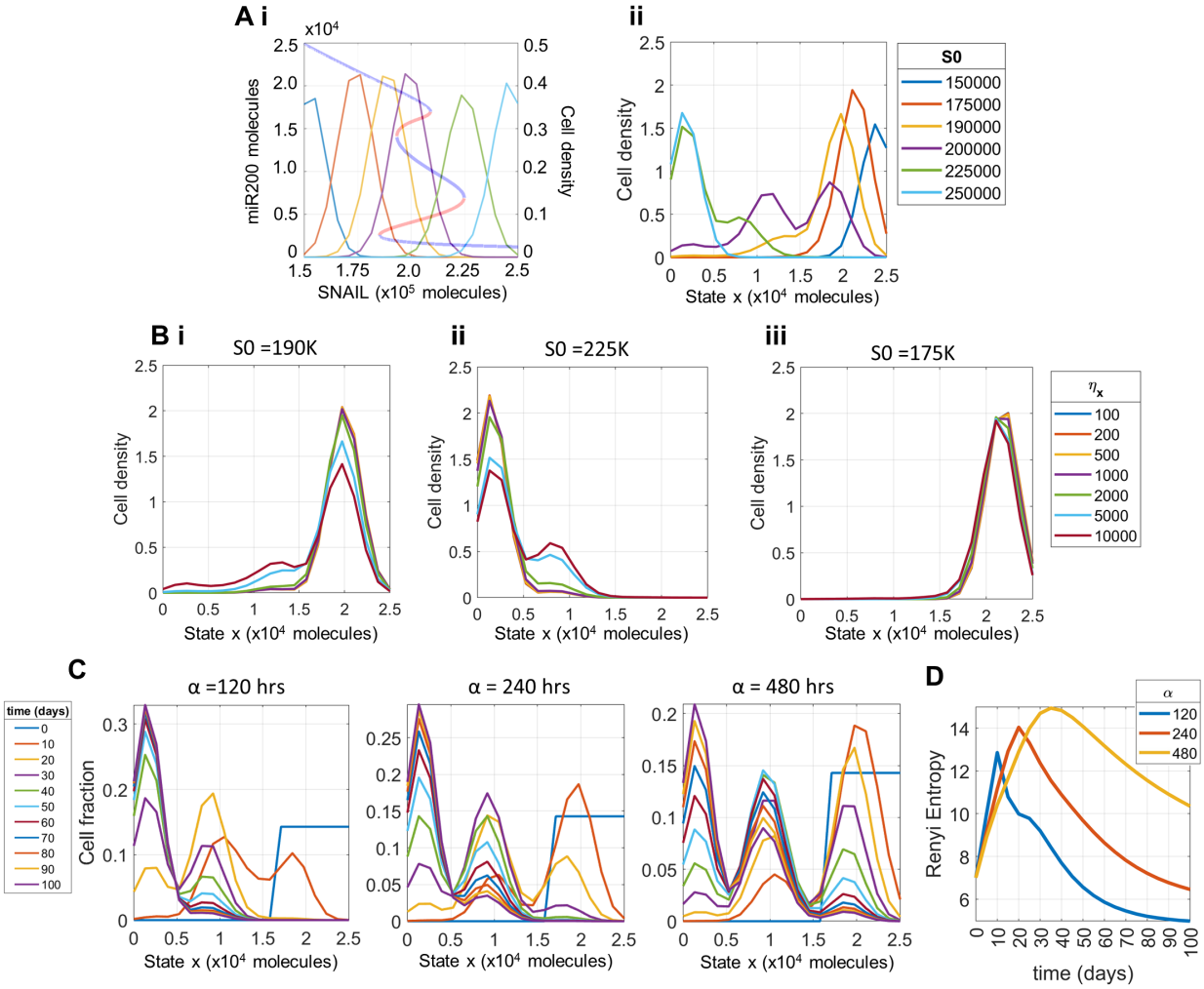

Figure 3: Simulations of the advection-selection-mutation model (PDE (5)) with reduced function f_r. Empirically reduced single cell state variable x, showing that the reduced system captures the full EMT network dynamical response; and the combined influence of biochemical noise with input signal dynamical response in modulating the phenotypic heterogeneity of the cell population. Ai) Bifurcation diagram showing stable (blue) and unstable (red) levels of miR200 (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 η_x for SNAIL distributions with mean $S_0 = 190K$ (Bi) and 225K (Bii). For $S_0 = 175K$ (Bii), 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 'α'). D) Time evolution of population's heterogeneity (measured by Renyi entropy) for different α parameter values. Parameters used to generate plots, unless stated otherwise, are $\alpha = 120hrs$, $\eta_x = 1000$, ini pop Epi (see Figure 4 A, i), time= 100 days, $S_0 = 225K$ molecules, and per-capita growth rate r is constant across phenotypes, given by $r = 0.0182/hr$.

cells. In practice, this means considering three different possible piecewise-constant functions r, as illustrated in Figure 4 A ii. Across all these cases, the E state divides at either an equal or a faster rate than hybrid E/M 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 hybrid E/M and/or M cells [27, 43].

First, we investigate how the phenotypic composition of the population changes with different growth rate scenarios (Figure 4, Figure S5). For uniformly distributed cells in all three E (epi), hybrid E/M (hyb), and M (mes) states (Figure 4 A, i) and SNAIL distribution with mean $S_0 = 190K$ or 200K molecules that predominantly enables an E state with/without hybrid E/M state, the reduced growth of M cells has very slight effects on phenotypic composition over the time course, as expected (Figure 4 B, i). The initial peak in hybrid cell fraction for the 'r2' growth scenario is the combined effect of M to hybrid E/M state transition and a relatively higher growth rate of hybrid E/M cells. However, when hybrid E/M 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 (Figure 4 B, i - $S_0 = 200K$). For the input SNAIL distribution with a mean value S_0 of 225K molecules that majorly supports hybrid E/M and M phenotypes, in the case 'r2', the growth advantage provided to hybrid E/M cells enables their dominance in the population on the long run, when compared to the growth scenarios of 'r1' or 'r3' where both E/M and M cells proliferate at equal rates (Figure 4 Bi, $S_0 = 225K$). The initial peaks in hybrid E/M fractions are combined effects of E to hybrid E/M transitions with either growth similarity or advantage of hybrid E/M cells over M cells.

The overall change in phenotypic composition can be calculated using Renyi entropy as a heterogeneity score (Figure 4 B, 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 look at the effects of increasing level of epigenetic noise level η_x in cell state x on phenotypic composition laid down by growth rate differences (Figure S6 A-D). For $S_0 = 190K$ and 200K, where hybrid E/M state is less dominant than the E state (Figure 4 B), increasing the noise levels (from $\eta_x = 1000$ to $\eta_x = 5000$) in state x causes more cell-state transitions, raising the frequency of hybrid E/M phenotype in population for all growth scenarios (Figure S6 B). However, as for $S_0 = 225K$ and growth scenario 'r2' where hybrid phenotype is the dominant state, increasing the noise η_x level (from $\eta_x = 1000$ to $\eta_x = 5000$) raises M fractions in the population even though M cells are dividing slowly (Figure 4 B,i; Figure S6 B-C). Overall, we observe an increase in population heterogeneity with higher noise levels in state variable 'x', irrespective of the growth scenario (Figure 4 Bii, Figure S6 D).

After observing changes in phenotypic composition for different growth scenarios, we move on to see how total cell population grows for combinations of initial conditions and growth scenarios. We consider six different conditions – isolated E, hybrid E/M, and M population, uniform mixture of either E and M or E, E/M 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 hybrid E/M cells (growth scenario 'r2'), we see that an initially mesenchymal population has slower population growth compared to other initial conditions. Similarly, with both hybrid E/M and M cells dividing slower than E (growth scenario 'r3'), initial conditions having component of either E/M or M cells divide slower than isolated (pure) E cells. Also, as transition from hybrid E/M to E is much more probable than M to E transitions, presence of hybrid cells in the populations increases the overall growth rate (Figure 4 C compare 'hyb vs mes', and 'epi mes' vs 'epi hyb mes' initial conditions for $S_0 = 200K$, 'r3' growth scenario). Further, increasing the level of stochastic noise (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 (Figure S6 E).

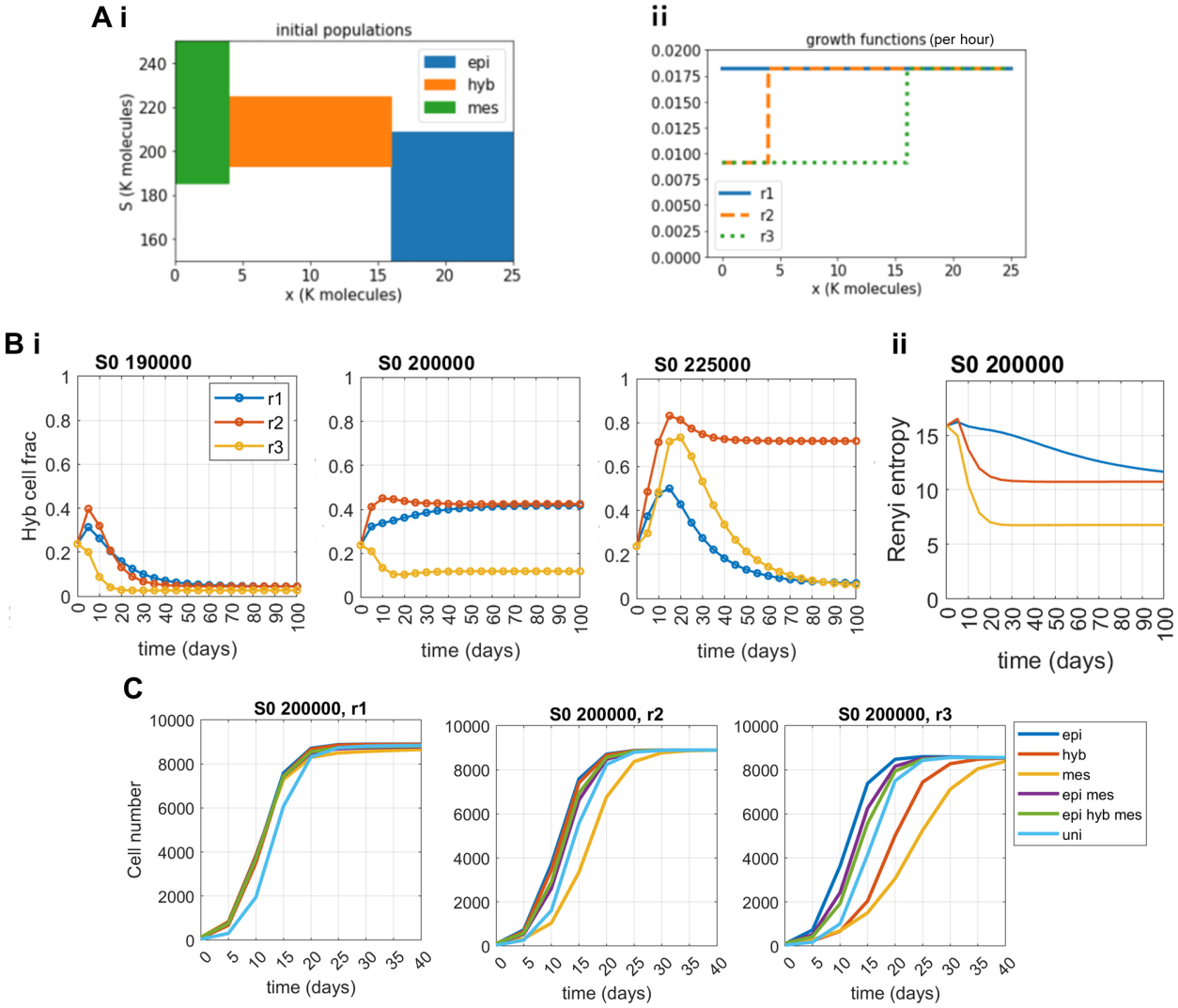

Figure 4: Effect of growth rate differences among E, hybrid E/M and M phenotypes on population's heterogeneity and overall growth dynamics. Ai) Support of the different initial conditions: In each case, the initial condition is uniformly distributed on its support, and such that the total initial population equals 100 cells. Aii) Profile of different growth functions, expressed per hour): 'r1': All three phenotypes divide at same rate; 'r2': E and E/M divide at equal rates, while M divide at half the rate of E cells; and 'r3': Both E/M 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 Renyi 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 colored domains of A1), and 'uni' to a uniform population on the whole rectangle $[0, 25K] \times [150K, 250K]$. For panel 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 $\alpha = 120hrs$, $\eta_x = 1000$, and the per-capita growth rate (r) of the epithelial phenotype is $0.0182/hr$.

2.5 Heterogeneity-dependent growth explains the faster tumour growth with highly heterogeneous parental and its subclones along E-M phenotypic axis

Experimental data from orthotopic implantation of SUM149 cells and its subclones with varying degree of E, M and hybrid E/M heterogeneity levels suggest that the parental cell line and subclones with high levels of E-M heterogeneity enabled the fastest tumour growth in mice [5]. Further, despite starting from varying levels of E-M heterogeneity in the orthotopic injected cell populations, all tumours have relatively higher E-M heterogeneity levels when mice were sacrificed [5]. These observations together indicated a plausible relationship between tumour growth rate and its heterogeneity. Thus, to explain these experimental observations, we assumed in our model formalism that the population growth rate depends on its heterogeneity levels (measured by differential entropy, a continuous equivalent of Renyi entropy). We specifically considered two functional relations between population growth rate and heterogeneity – 1) Linear relation, and 2) Sigmoidal relation (using Hill function with different threshold levels) (Figure 5A).

To computationally track the population dynamics, we kept the S levels of cells to be invariant such that the stability landscape along cell state 'x' remained the same for all simulation times (Figure 5 B). We started our simulations with five different populations (distributed along state x) – Pure E (ep), pure M (mes) and pure hybrid E/M (hyb) populations, mixed E and M population (ep_mes), and uniformly distributed population in E, M and E/M states (Figure 5 C). the difference in the heterogeneity among each starting population can be seen at zero-hour time point in Figure 5D. Further, as the cells divide and undergo cell-state transitions, the population heterogeneity value changes with time. All five different initial population distributions asymptotically reach saturating levels of heterogeneity, which is jointly determined by the stability landscape (Figure 5 B) and noise levels (η_x). Given that the population growth rate depends on its heterogeneity, we noticed that populations with higher E-M heterogeneity to start with (hyb, ep_mes, and unif; Figure 5D) show faster growth than other population with lower levels of E-M heterogeneity at the start (ep, and mes; Figure 5 D, E). The differences between the growth curves corresponding to different initial populations are much more pronounced for a linear than sigmoidal relationship between growth and heterogeneity because of its large variance in growth rates within the variability range of population's heterogeneity (Figure 5 A, D, E and Figure S7). Overall, by assuming the population growth rate to be dependent on its heterogeneity, we could recapitulate qualitatively the experimentally observed differences in the tumour growth dynamics in mice [5].

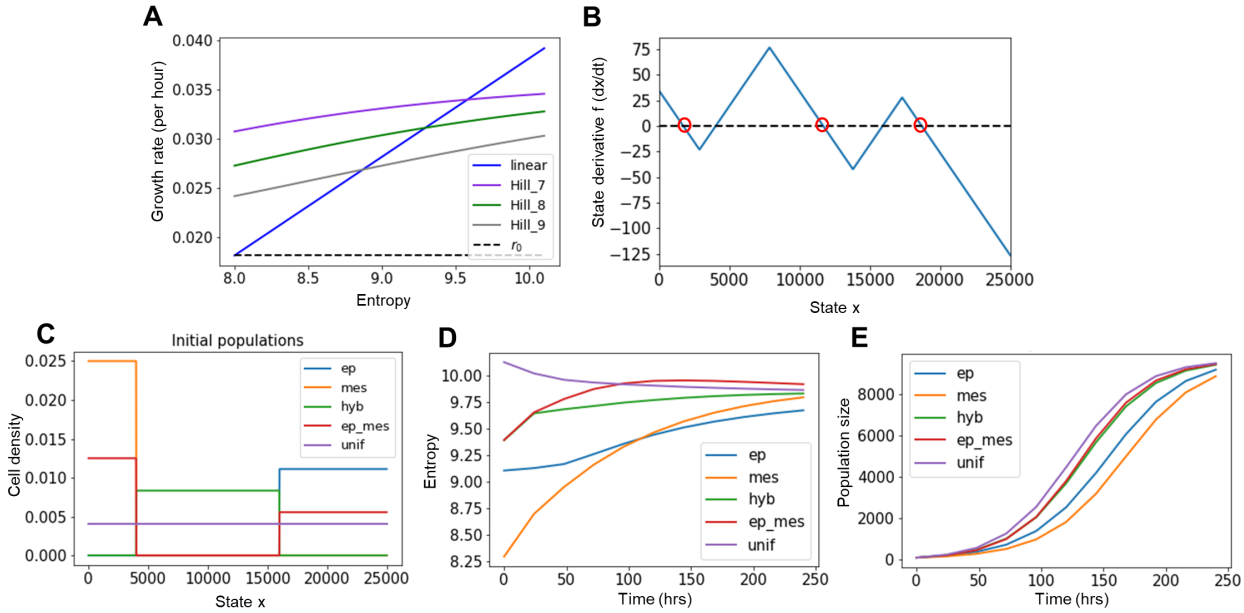

Figure 5: Heterogeneity-dependent growth of cells captures the in-vivo tumour growth dynamics observed in Figure 3A of [5]. A) Cellular division/growth rate as a function of population heterogeneity. Two functional dependences are considered – 1) Hills relation with increasing threshold value, and 2) Linear relation. More information about the functional relation is present in the Methods section. B) Reduced EMT state function f_r (underlying the reduced ODE) with three stable states (encircled) for constant S level of $200K$. C) Different starting population to simulate population dynamics. D) Time evolution of the population heterogeneity (entropy) for different initial distributions of cells in E, E/M and M phenotypes when growth rates of cell have linear relation with entropy (linear in Panel A). E) Time evolution of total population size for simulation in panel D. Uniform (unif), Epithelial and Mesenchymal (ep_mes), and Hybrid (hyb) starting populations grow faster because of their highly heterogeneous population at all time points.

3 Discussion

Several intracellular and intercellular regulatory and stochastic process 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 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 dynamic 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 and, 2) heterogeneity of growth rates among distinct subpopulations.

Our analysis is based on a minimalistic three node EMT regulatory network with a characteristic phenotypic landscape (Figure 2 A) [40]. Our choice of a sufficiently simple such network was based on the following criteria – a) making the analysis computationally tractable, and b) integrating multiple processes together – cell division, death, cell-state transition and intracellular regulatory dynamics. However, many more complex EMT regulatory networks have been identified over the past decade [18]. Further, the minimalistic EMT model considered here does not have a regulatory term for the input signal, so we assumed the input to have negative feedback onto itself and that its levels in the population is distributed around

the mean S_0. The 'α' and S_0 in input 'S' dynamics (Equation (6)) can be considered as the inverse of the birth rate and ratio of birth to death rates of S molecules, respectively. Therefore, the parameter 'α' sets maximum rate at which the population reaches its higher mean S_0 levels in the event of perturbation of the population distribution (Figure 3 C). With each molecule having its own birth and death kinetics, we see that the cellular memory can span across generations by inheriting the levels of a particular protein, as seen experimentally (Figure 3, [44, 45]).

To further reduce the long computation of the full EMT network when simultaneously considering cell division, death and mutation terms, we reduced the existing two-variable one input signal EMT state dynamical model to single-variable one input signal EMT model. The parameters of the reduced one-dimensional state derivative function (Figure S2 A) were set by minimising the error of its resulting dynamics with the evolution of cell state while two-dimensional state derivative (full EMT model) was considered. 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 [32, 35, 46]. Although, 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 mutation kernel, which we then assumed to be normally distributed about 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 process on cellular heterogeneity [31, 47, 33], and the model framework developed here can be easily used to account for other stochastic factors.

Our framework encompasses considering a diffusion term as has been done elsewhere. Indeed, a second-order diffusion term is the PDE counterpart of adding a Brownian motion to the considered ODE as in [30]. The corresponding diffusion term can be recovered by a suitable scaling of our integral mutation term as explained in [48], which we have not done in the present work. Additionally, by approximating 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 (Figure 3 B). The main equation studied in this article (5) is in line with Cell Population Balance models usually employed for heterogeneous cell populations [49, 50, 34, 36]. 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 mutation term are both non-local terms, since the former depends on the population size ρ, 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 [39]. Compared with other methods used for the same type of problems, such as finite element methods [51] or finite difference methods [52], 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 (2) to the full model with growth and epimutations (5), and then to the entropy-dependent growth equation (10)). 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 mutation terms, particle methods are not asymptotic-preserving schemes [38], which means that the particle approximation (8) 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.

By drawing a simplistic relationship from the experimental data between in vivo growth and tumour

heterogeneity, we were able to explain trends in the tumour growth dynamics as seen (Figure 5). However, we understand that various other factors such as feedback loops formed by interactions of tumor cells with extra-cellular matrix (ECM) and/or other stromal cells can alter heterogeneity patterns as well. For example, cells undergoing EMT can secrete LOXL2 that increases collagen crosslinking in ECM, and the ECM density as a well as stiffness can induce EMT.

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 with an efficient PDE numerical integration scheme recently developed by some of us [39].

4 Methods

Introduction to phenotype-structured PDE models

The state of a given cell is described by a time-dependent vector $y(t) \in \mathbb{R}^n$: for a given $i \in \{1, \dots d\}$, $y_i(t)$ represents the concentration of some protein i inside the cytoplasm of the cell at time t. In the context of EMT, the y_i's can represent the level of several EMT markers such as miR-200 (μ_{200}), ZEB (Z) or SNAIL (S). The time-evolution of the cell state, for a single cell, is modelled by an Ordinary Differential Equation (ODE) of the form

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

where $\dot{y}$ denotes the derivative with respect to time, and $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is a function describing the interactions between different molecules, which will be called *advection function* throughout. Such an ODE is solved once complemented with an initial condition $y(0) = y^0$ where $y^0 \in \mathbb{R}^n$.

A population composed of many cells can be described by means of a density function $u(t, y)$, which represents the number of cells of state $y \in \mathbb{R}^n$ at time $t \geq 0$. In other words, the number of cells whose state lies in some set $\mathcal{E} \subset \mathbb{R}^n$ at a given time t is given by $\int_{\mathcal{E}} u(t, y)dy$. The time-evolution of $u(t, y)$ associated to (1) is then given by a PDE, the so-called *advection equation*, namely

$$\partial_t u(t, y) + \nabla \cdot (f(t, y)u(t, y)) = 0, \tag{2}$$

where ∂_t denotes partial derivation with respect to time, and $\nabla\cdot$ denotes the (partial) divergence operator with respect to the variable y. Such a PDE is solved once complemented with an initial condition $u(0, \cdot) = u^0$ with $u^0 : \mathbb{R}^n \to \mathbb{R}$.

Simulating hysteresis and epigenetic regulation

We begin by considering the advection function associated with a minimal gene regulatory network developed in [40], which writes,

$$\begin{cases} \dot{\mu}_{200} &= g_{\mu_{200}} H_{Z,\mu_{200}}(Z) H_{S,\mu_{200}}(S) - g_{m_Z} H_{Z,m_Z}(Z) H_{S,m_Z}(S) Q(\mu_{200}) - k_{\mu_{200}} \mu_{200} \\ \dot{Z} &= g_Z g_{m_Z} H_{Z,m_Z}(Z) H_{S,m_Z}(S) P(\mu_{200}) - k_Z Z \end{cases}. \tag{3}$$

We denote by $F = F(\mu_{200}, Z, S)$ 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 μ_{200}). The variable S represents a third molecule, SNAIL, which is seen in our case as an external signal characterising the extracellular environment. All the parameters of this model are given in Appendix B. Note that this model falls into the framework introduced in equation (1), 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 displayed in Figure 2 A depicts the different possible stable states, each characterised by specific ranges of miR-200 levels (solid lines) for increasing levels of SNAIL, resulting from the network dynamics. As a cell undergoes EMT (i.e. as SNAIL levels increase), it switches from high to intermediate to low levels of miR-200, which corresponds to the E, E/M and M states respectively. However,

during MET, the cell switches directly from low (M) to high (E) miR-200 levels without passing through the hybrid E/M state, thus displaying hysteresis.

Homogeneous population, (Figure 2 C i). To reproduce the hysteretic behaviour of EMP, *i.e.* asymmetry in EMT and MET trajectories, we first consider the advection equation (2), 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 $(10000, 160K)$, as represented in Figure 2B (black).

Heterogeneous population, (Figure 2 C ii). 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 incorporate 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 $S \in (5000, 10000]$. As initial condition, we consider a population homogeneously distributed in the molecules ZEB and miR-200, but with heterogeneous levels of SNAIL distributed according to a Gaussian. In other words, we let $u^0(\mu_{200}, Z, S) = \frac{1}{\sigma} G(\frac{S - S_0}{\sigma})$, where $S_0 = 160K$, $\sigma = 20K$ and G denotes the Gaussian function.

Epigenetic regulation, (Figure 2 D). Lastly, we run simulations similar to those carried out for a homogeneous population, but with a modified advection function which allows to account for epigenetic regulation. We incorporate 'Z_0' into the structure variable, which then writes $y = (\mu_{200}, Z, Z_0, S)$. This new parameter represents the ZEB threshold for inhibiting miR-200. The considered advection function is that associated to the ODE

$$\begin{cases} \dot{\mu}_{200} &= g_{\mu_{200}} H_{Z,\mu_{200}}(Z_0, Z) H_{S,\mu_{200}}(S) - g_{m_Z} H_{Z,m_Z}(Z) H_{S,m_Z}(S) Q(\mu_{200}) - k_{\mu_{200}} \mu_{200} \\ \dot{Z} &= g_Z g_{m_Z} H_{Z,m_Z}(Z) H_{S,m_Z}(S) P(\mu_{200}) - k_Z Z \\ \dot{Z}_0 &= \frac{1}{\beta(t)} \left(Z^0_{\mu_{200}} - Z_0 - \alpha Z \right), \end{cases} \tag{4}$$

whose parameters are detailed in Appendix B. Denoting F_e the right-hand side of this ODE, the corresponding PDE model (2) 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 (Figure 2 D i), S is the piecewise-linear function which connects the points $(0, 100K), (1200, 240K)$ and $(2400, 100K)$, as represented in Figure 2 B (blue curve), and in a second simulation (Figure 2 D ii), the piecewise-linear function which connects the points $(0, 100K), (1200, 240K)$ and $(2400, 240K)$ and $(3600, 100K)$ (red curve in Figure 2 B).

The simulations for these three models are performed with a particle method detailed later on in this Methods section.

Reducing the dimensions of the structure variable.

When incorporating growth and mutations into the model, computation times become prohibitive. To ease the burden, we use dimension reduction by further simplifying the advection function: we consider the state $y = (x, S)$ rather than (μ_{200}, Z, S), where x is roughly equivalent to μ_{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 Figure 2 A, 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 assume 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 Appendix C.

Considering growth and epimutations.

The full PDE model incorporating growth and epimutations writes

$$\begin{cases} \partial_t u(t,y) + \nabla \cdot (f(y)u(t,y)) = (r(y) - d(y)\rho(t))\, u(t,y) \\ \qquad\qquad + \int_{\mathbb{R}^2} M(y,z)u(t,z)dz - \int_{\mathbb{R}^2} M(z,y)dz\, u(t,y) \\ \rho(t) = \int_{\mathbb{R}^2} u(t,y)dy \end{cases}, \tag{5}$$

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

$$f_S(S) = \delta\big(1 - \frac{S}{S_0}\big), \tag{6}$$

  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.
- The term $(r(y) - d(y)\rho(t))\, u(t,y)$ represents the net growth of cells of state y, with two main contributions given by the intrinsic growth rate $r(y)$, and the death rate $d(y)\rho(t)$ proportional to the total population size $\rho(t)$. This corresponds to the so-called logistic model, accounting for the additional death rate due to competition for resources and space between cells. In all our simulations, the death rate is considered to be independent of the cell state (*i.e.* $d(y) = d \equiv 1.82 \times 10^{-7}$cell/hr, as in [24]).
- The last two terms represent cell mutations, and can again be decomposed into two terms. The term $\int_{\mathbb{R}^2} M(y,z)u(t,z)dz$ represents the mutation of cells of any type z into cells of type y, occurring with a rate $M(y,z)$. The term $\int_{\mathbb{R}^2} M(z,y)dz\, u(t,y)$ represents the mutation of cells originally of type y into cells of any other type z, with mutation rate $M(z,y)$. The mutation function M is taken to be $M(y,z) = r(z)P(y-z)$, meaning that mutations are considered to occur at cell division. Here, $P(y) = P(x,S) := \frac{1}{\eta_x \eta_S} G(\frac{x}{\eta_x}) G(\frac{S}{\eta_S})$, where G is the Gaussian function. Variables η_x and η_S are the standard deviations for x and S respectively.

Numerical method

For numerical simulations, 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

$$\begin{cases} \partial_t u_{re}(t,y) + \nabla \cdot (f_{re}(y)u_{re}(t,y)) = (r_{re}(y) - d_{re}(y)\rho(t))\, u_{re}(t,y) \\ \qquad\qquad + \int_{\mathbb{R}^2} M_{re}(y,z)u(t,z)dz - \int_{\mathbb{R}^2} M_{re}(z,y)dz\, u(t,y) \\ \rho(t) = \int_{\mathbb{R}^2} u_{re}(t,y)dy \end{cases}, \tag{7}$$

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 $u^0_{re}(x,S) := ACu^0(Ax+B, CS+D)$.

Thus, an approximation for u_{re} provides an approximation for u. We apply a particle method in order to approximate u_{re} at different time steps ($0 < T_1 < ... < T_K$, specified in each figure), applying a particle method introduced in [39] to deal with a category of models to which (5) belongs. For a proof that the numerical scheme does successfully approximate the solutions of (5), we refer to [39], while an introduction to particle methods can be found in [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 $i \in \{1, ..., N^2\}$, we solve the ODE

$$\begin{cases} \dot{y}_i = f(y_i) \\ \dot{w}_i = \nabla \cdot f(y_i)\, w_i \\ \dot{v}_i = \left(r(y_i) - d(y_i) \sum\limits_{j=1}^{N^2} v_j \right) v_i + w_i \sum\limits_{j=1}^{N^2} M(y_i, y_j) v_j - \sum\limits_{j=1}^{N^2} M(y_j, y_i) v_i \end{cases}, \tag{8}$$

with initial conditions $y_i(0) = y_i^0$, $w_i(0) = \frac{1}{N^2}$ and $v_i(0) = \frac{u_{re}^0(y_i^0)}{N^2}$, on the first time interval ($[0, T_1]$). The solution of this ODE is called *Particle approximation of* (5). To solve (8) 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 [53].

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

$$u_\varepsilon^N(T_1, x) = \sum_{i=1}^{N^2} v_i(T_1)\, G_\varepsilon\big(y - y_i(T_1)\big), \tag{9}$$

with $G_\varepsilon(x, S) := \frac{1}{\varepsilon^2} G(\frac{x}{\varepsilon}) G(\frac{S}{\varepsilon})$, with G the Gaussian function, and $\varepsilon = \left(\frac{1}{N^2}\right)^\gamma$, 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 mutations), and by comparing with other values of γ. 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 $u^N(T_k, \cdot)$.

Entropy-dependent growth.

In Figure 5, we consider a model for which the growth function r depends on population heterogeneity. For simplicity, we assume that the SNAIL level is constant within the population, and does not change over time. Thus, we use the simplified advection function f_r with $S \equiv 200K$, which allows for the existence of the three phenotypes as shown on the bifurcation diagram (Figure 2 A). We measure heterogeneity within the population by means of the entropy $-\int_{\mathbb{R}} \frac{u(t,y)}{\rho(t)} \ln(\frac{u(t,y)}{\rho(t)}) dy$. The model writes

$$\begin{cases} \partial_t u(t,x) + \nabla \cdot (f(x) u(t,x)) = (r(x, E(t)) - d(x, E(t))\rho(t))\, u(t,x) \\ + \int_{\mathbb{R}} M(x, y, E(t)) u(t,y) dy - \int_{\mathbb{R}} M(y, x, E(t)) dy\, u(t,x) \\ \rho(t) = \int_{\mathbb{R}} u(t,y) dy \\ E(t) = -\int_{\mathbb{R}} \frac{u(t,y)}{\rho(t)} \ln\left(\frac{u(t,y)}{\rho(t)}\right) dy \end{cases}, \tag{10}$$

with $f(x) := f_r(x, 200K)$, and $M(x,y) = \frac{1}{\eta_x} G\big(\frac{x-y}{\eta_x}\big) r(y, E)$ with G the Gaussian function and $\eta_x = 4000$, $d(x, E) = r(x, E)/10K$. We take four different values for the growth function 'r': shifted Hill functions of the shape $r_0 \frac{\theta^6 + 2E^6}{\theta^6 + E^6}$, with $\theta \in \{7, 8, 9\}$, $r_0 = 0.0182$ ($\theta = 9$ in Figure 5), and a linear function $r_0 + 0.01(I - 8)$. These four functions are represented in Figure 5 C.

We use the same numerical scheme as for the previous equation (in one dimension), but with the following ODE

$$\begin{cases} \dot{x}_i = f(x_i) \\ \dot{w}_i = \nabla \cdot f(x_i)\, w_i \\ \dot{v}_i = \big(r(x_i, \bar{E}) - d(y_i, \bar{E})\bar{\rho}\big)\, v_i + w_i \sum\limits_{j=1}^{N} M(y_i, y_j, \bar{E}) v_j - \sum\limits_{j=1}^{N} M(y_j, y_i, \bar{E}) v_i \\ \bar{\rho} = \sum\limits_{j=1}^{N} v_j, \quad \bar{E} = -\sum\limits_{j=1}^{N} \frac{v_j}{\bar{\rho}} \ln(\frac{v_j}{w_j \bar{\rho}}) \end{cases}, \tag{11}$$

with parameters $N = 50$ and $\varepsilon = \left(\frac{1}{N}\right)^{0.8}$.

Funding. This work was supported by the Raman-Charpak Fellowship 2022 awarded to J.G., funding a two-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.

A Supplementary figures

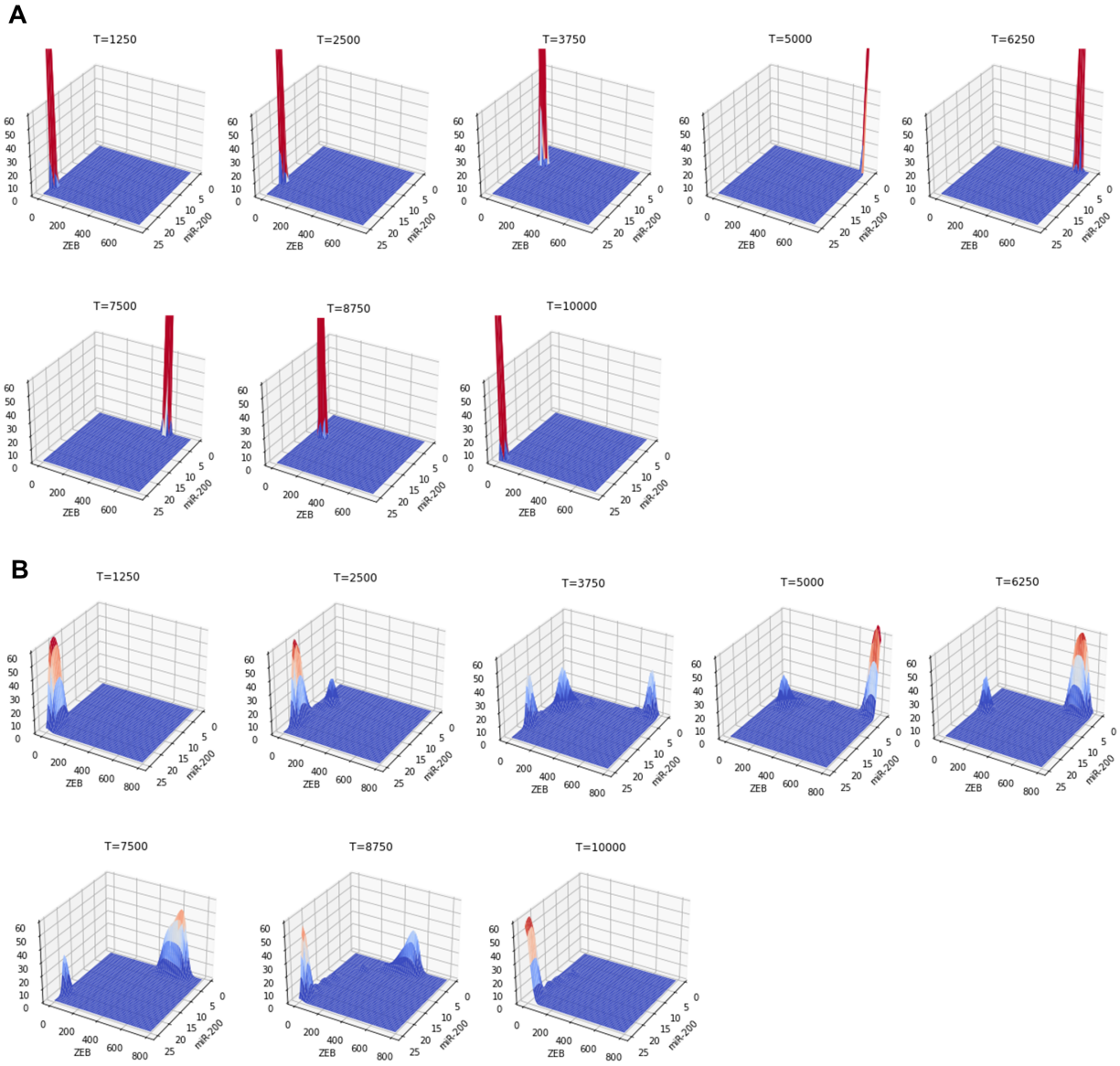

Figure S1: Hysteresis (non-symmetric transition) in cell density (z-axis) along the two cell state variables miR200 and ZEB during one cycle of EMT and MET caused by increasing and decreasing levels of input SNAIL levels (Figure 2 B blue curve) for A) homogeneous and B) heterogeneous cell population.

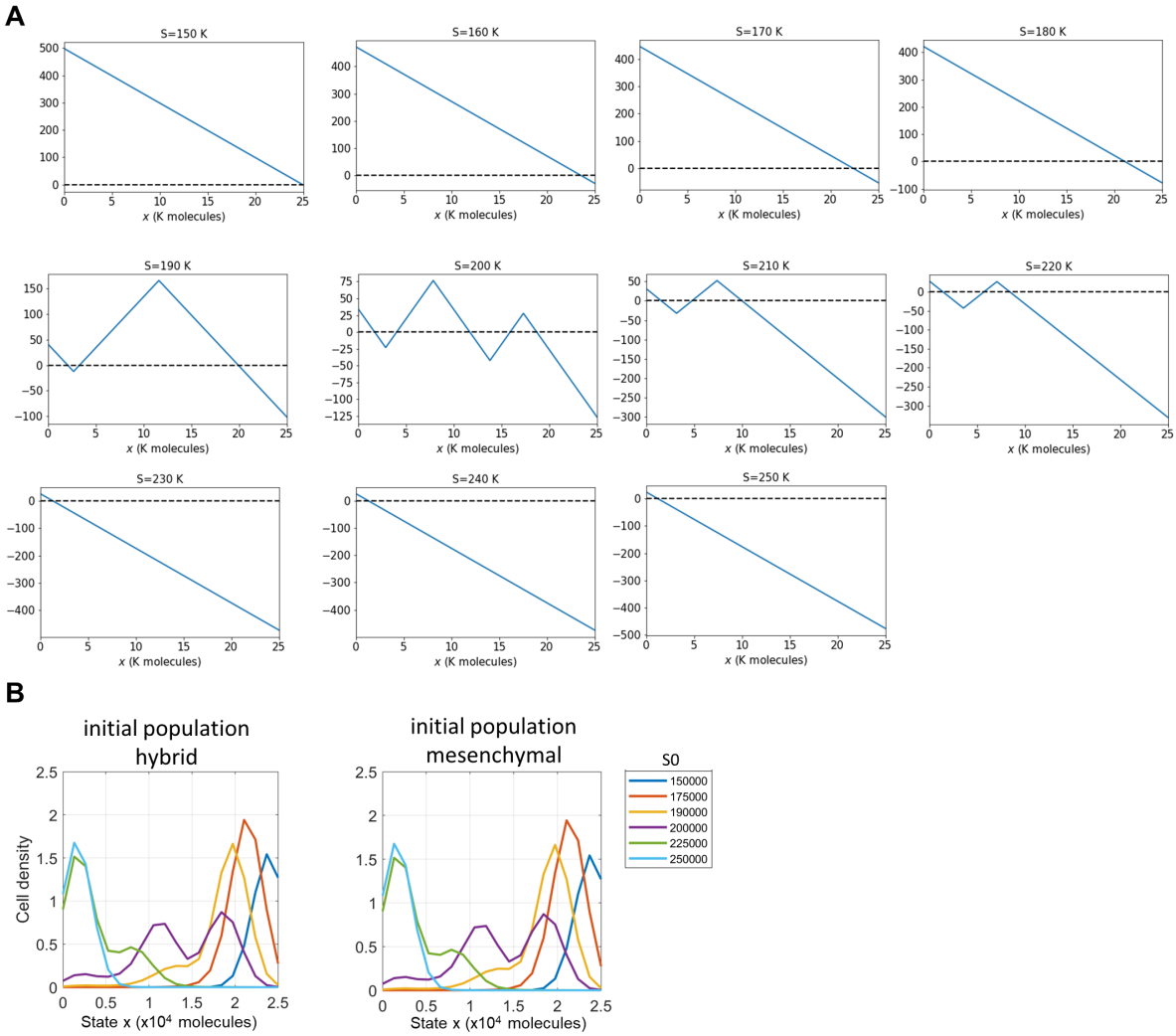

Figure S2: Reduced function f_r underlying the ODE with single variable cell state x and single input SNAIL; and distribution of the cell population for different distributions of SNAIL shown in Figure 2 Ai. A) Function $x \mapsto f_r(x, S)$, showing the existence of mono-, bi-, and tri-stable states for varying levels of SNAIL inputs S. B) Cell population distribution at the end point (100 days) of simulations started with hybrid and mesenchymal populations for increasing levels of input signal SNAIL's mean (S_0) levels. Parameters used to generate the above plots, unless stated otherwise, are $\alpha = 120$ hrs, $\eta_x = 1000$, ini pop Epi, time point 100 days, $S_0 = 200K$ molecules, and the per-capita growth rate (r) is constant $= 0.0182\backslash hr$.

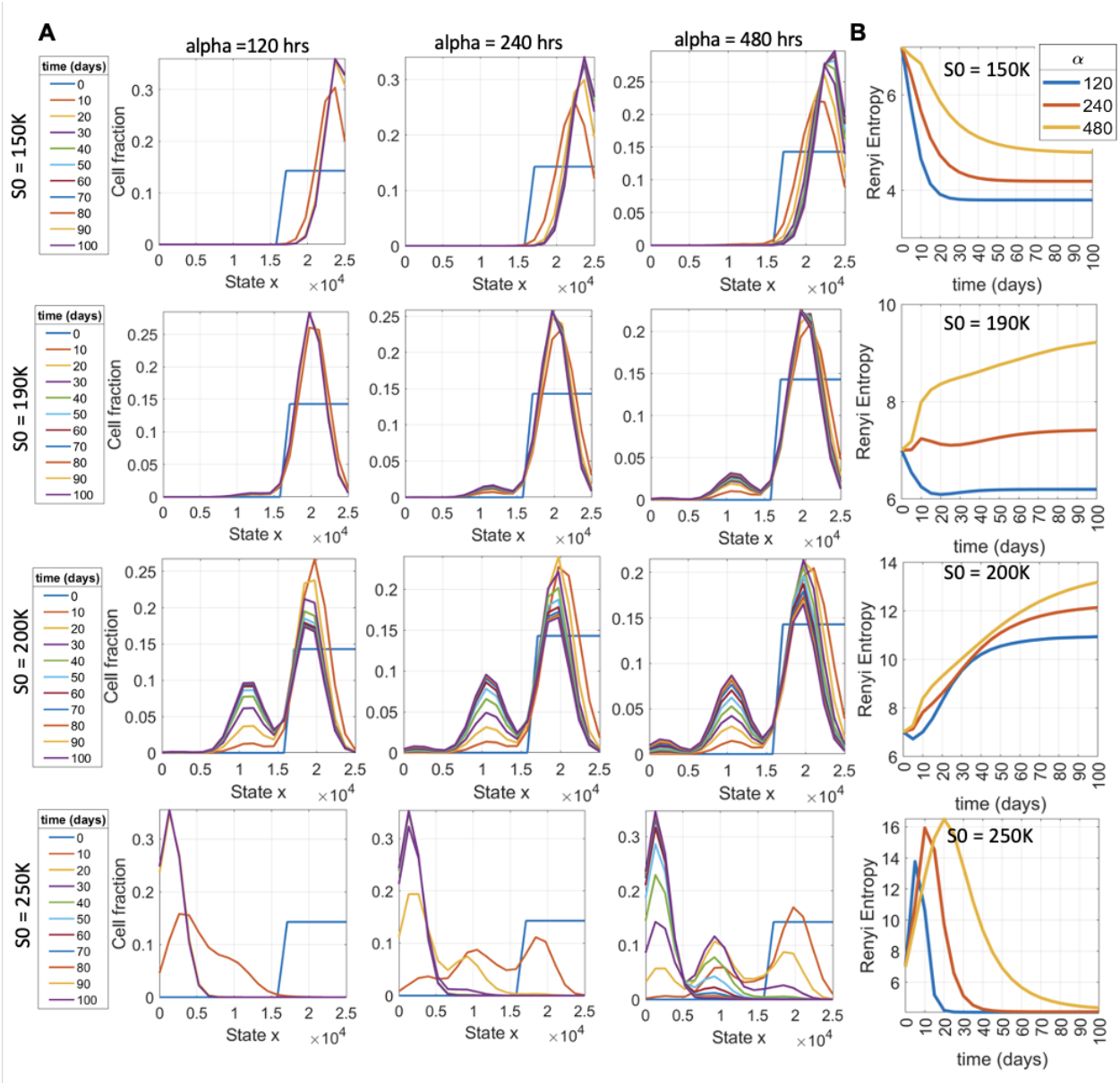

Figure S3: Population dynamics and changes in population heterogeneity (measured using Renyi Entropy) with time for several combinations of S_0 and α parameters. The intermediate S_0 values of 200K molecules where a cell can either attain a stable epithelial, hybrid, and mesenchymal, has highest heterogeneity score. Further, as we go towards smaller or larger values of S_0 where only epithelial, and mesenchymal states are possible, respectively, the population has least heterogeneity. And, by increasing the residence time of cells in a state x, cells get enough time to populate the cell state they reside in, and thereby, contribute to a significant fraction in the overall population. For the above plots, the initial condition population is epithelial, $\eta_x = 1000$, and per-capita growth rate r is constant $= 0.0182/hr$.

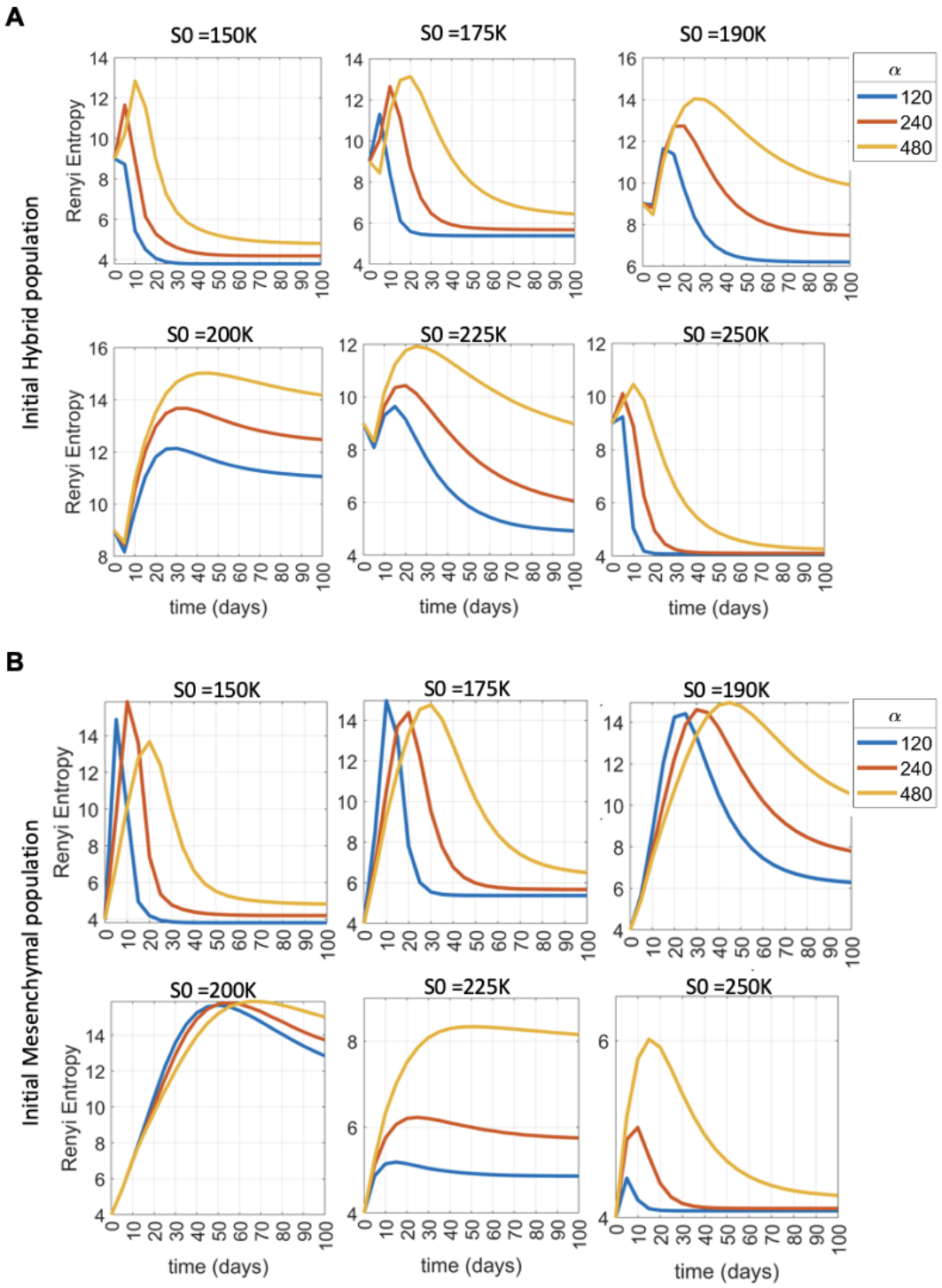

Figure S4: Changes in population heterogeneity (measured using Renyi Entropy) with time for several combinations of S_0 and α values while starting with a population of A) Hybrid cells, and B) Mesenchymal cells. For the above plots, and $\eta_x = 1000$, and per-capita growth rate r of all subpopulation $= 0.0182$/hr.

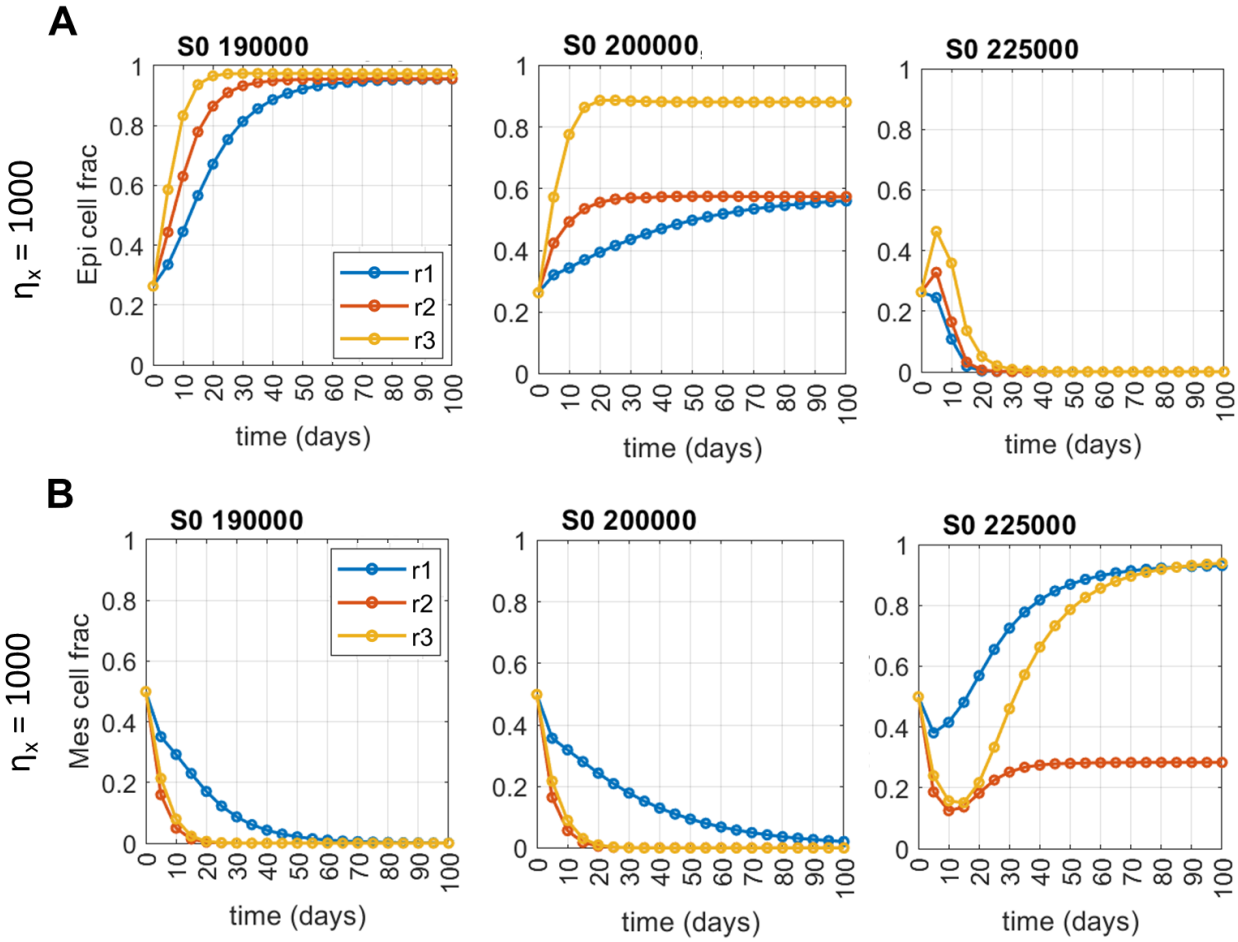

Figure S5: Effect of growth rate differences among E, hybrid E/M and M phenotypes on population's heterogeneity. Temporal changes in E cell fraction (panel A) and M cell fraction (panel B) in the population for different growth scenarios among phenotypes – 'r1': All three phenotypes divide at same rate; 'r2': E and E/M divide at equal rates, while M divide at half the rate of E cells; and 'r3': Both E/M and M divide at equal but half the rate of E cells. For the above results, the initial condition is uniformly distributed in E, hybrid E/M and M state. The input SNAIL mean S_0 levels used are mentioned for all the individual plots. Other parameters used to generate plots are $\alpha = 120$ hrs, $\eta_x = 1000$, and per-capita growth rate r of epithelial phenotype is 0.0182/hr.

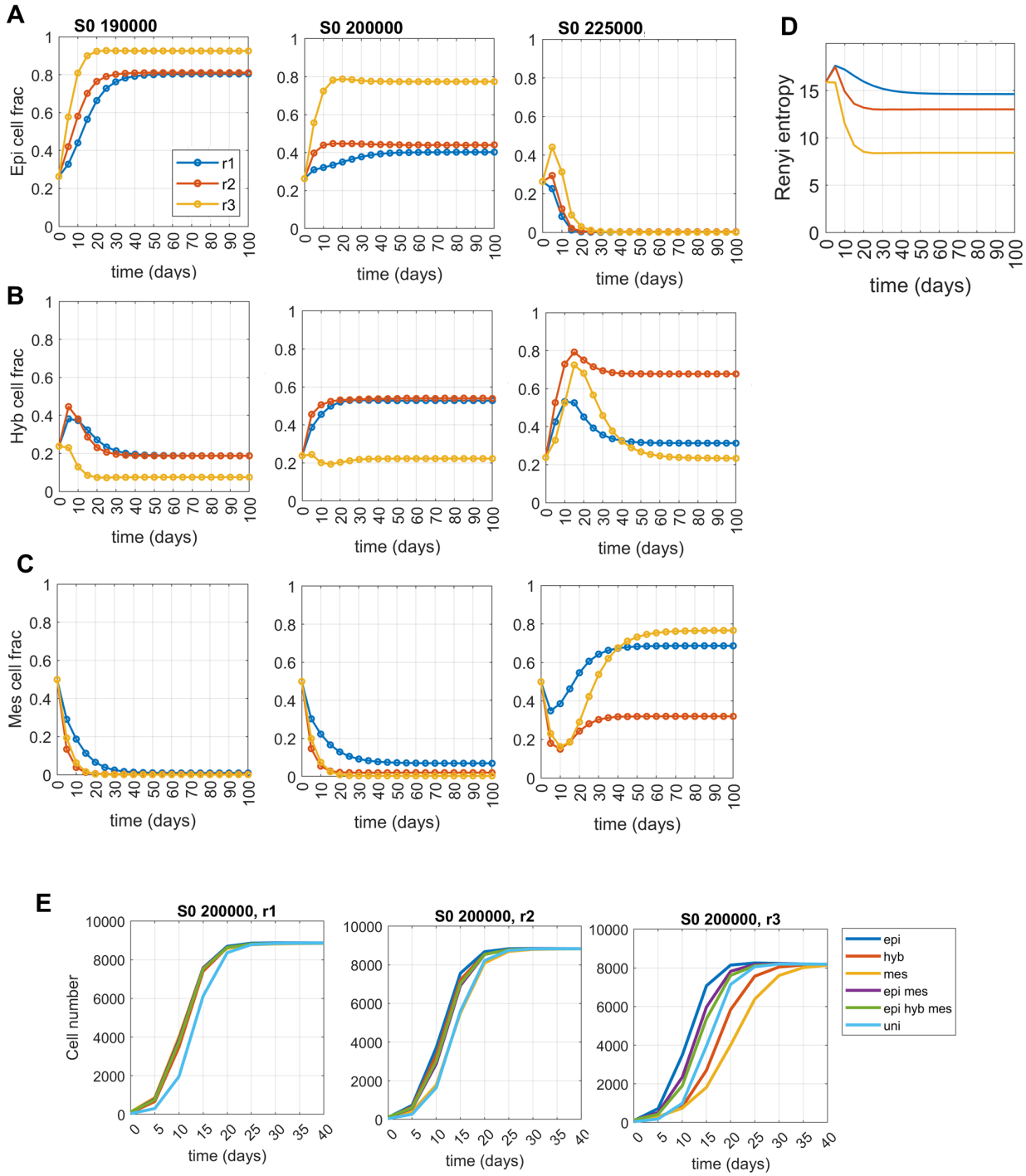

Figure S6: Effect of growth rate differences among E, hybrid E/M and M phenotypes on population's heterogeneity and overall growth dynamics. A-C) Temporal changes in E, hybrid E/M, and M cell fraction in the population for different growth scenarios among phenotypes – 'r1': All three phenotypes divide at same rate; 'r2': E and E/M divide at equal rates, while M divide at half the rate of E cells; and 'r3': Both E/M and M divide at equal but half the rate of E cells. D) Changes in population's heterogeneity (measured by Renyi entropy) with time. E) Population growth dynamics for different combinations of growth scenarios and initial conditions. For panel A-D, the initial condition is uniformly distributed in E, hybrid E/M and M state. The input SNAIL mean S_0 level used are mentioned for all the individual plots. Other parameters used to generate plots are $\alpha = 120$ hrs, $\eta_x = 5000$, and per-capita growth rate r of epithelial phenotype is 0.0182/hr.

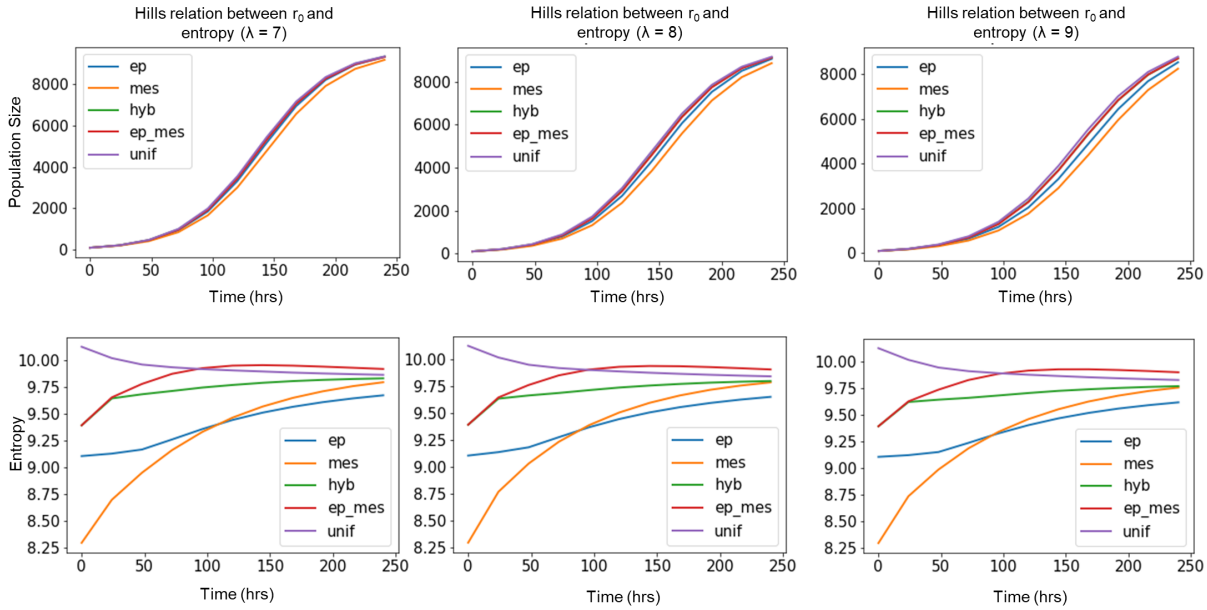

Figure S7: Temporal changes in total population size and population heterogeneity when growth rate of cell have sigmoidal relation with population heterogeneity (entropy). Parameters used to obtain above plot are $\alpha = 120$ hrs, $\eta_x = 1000$, and S levels in the population are set to 200K.

B Parameters for ODE (3) and (4)

We detail the parameters underlying ODE (3):

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

Functions $H_{Z,\mu_{200}}$, $H_{S,\mu_{200}}$, H_{Z,m_Z} and H_{Z,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 S1.

					Molecules		Molecules.Hour^{-1}		Hour^{-1}
n	6			μ_0	10K				
$n_{Z,\mu_{200}}$	3	$\lambda_{Z,\mu_{200}}$	0.1	$Z^0_{\mu_{200}}$	$220K$	$g_{\mu_{200}}$	$2.1K$	$k_{\mu_{200}}$	0.05
$n_{S,\mu_{200}}$	2	$\lambda_{S,\mu_{200}}$	0.1	$S^0_{\mu_{200}}$	$180K$	g_Z	$0.1K$	k_Z	0.1
n_{Z,m_Z}	2	λ_{Z,m_Z}	7.5	$Z^0_{m_Z}$	$25K$	g_{m_Z}	11	k_{m_Z}	0.5
n_{S,m_Z}	2	λ_{S,m_Z}	10	$S^0_{m_Z}$	$180K$				

Table S1: Parameters for the Hill functions.

The functions Y_μ, Y_m and L are defined by

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

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

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

Parameters for these three functions are given in Table S2.

i	0	1	2	3	4	5	6
l_i	1.0	0.6	0.3	0.1	0.05	0.05	0.05
γ_{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 S2: Parameters for the functions L, Y_μ and Y_m

The additional parameters of ODE (4) are $\alpha = 0.15$, and $\beta(t) = 240.$ when S is non-decreasing, and $\beta(t) = 720.$ when S is decreasing.

C 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 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 mensechymal phenotype), as can be seen on the bifurcation diagram (Figure 2 A).

The values of stable equilibrium points representing the three possible phenotypes slightly varyn 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

$$a_5S^5 + a_4S^4 + a_3S^3 + a_2S^2 + a_1S + a_0.$$

$P_{mes}(S), P_{hyb}(S)$ 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 S3.

	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	$3.430402e+06$	$-8.774602e+07$	$4.005964e+08$	$-6.573913e+09$	$3.1850617e+06$

Table S3: Coefficients of the polynomials P_{mes}, P_{u1}, P_{hyb}, P_{u2} and P_{ep}.

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 $x \leq 0.5(P_{mes}(S) + P_{u1}(S))$, $f_r(x, S) = -(x - P_{mes}(S))$
  - 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 look for the value of k that minimises the quantity

$$\int_0^T \int_{150K}^{250K} \int_0^{25K} \int_0^{800K} |x(t, S, x_0) - \mu(t, S, x_0, Z_0)| dt\, dS\, dx_0\, dZ_0, \tag{12}$$

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

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

and for all $Z_0 \in [0, 800K]$, $(\mu(\cdot, S, x_0, Z_0), Z(\cdot, S, x_0, Z_0))$ solves (3).

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

$$\frac{1}{N_T N_S N_x N_Z} \sum_{i,j,k,l} |x(t_i, S_j, x0_k) - \mu(t_i, S_j, x0_k, Z0_l)|,$$

with $N_S = N_x = N_Z = 20$, $N_T = 100$, and for all i, j, k, l, $t_i = i\frac{T}{N_T}$, $S_j = 150K + j\frac{100K}{N_S}$, $x0_k = k\frac{25K}{N_x}$, and $Z0_l = l\frac{800K}{N_Z}$, where $x(\cdot, S_j, x0_k)$ solves

$$\begin{cases} \dot{x}(t, S_j, x0_k) = f_r\left(x(t, S_j, x0_k)\right) \\ x(0, S_j, x0_k) = x0_k \end{cases},$$

$(\mu(\cdot, S_j, x0_k, Z0_l), Z(\cdot, S_j, x0_k, Z0_l))$ solves (3).

The value of the multiplicative constant which minimises (12) 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 it the value that we select. The obtained function is shown in Figure S2 for various values of S.

References

[1] Jacquemin V, Antoine M, Dumont JE, Dom G, Detours V, Maenhaut C. Dynamic Cancer Cell Heterogeneity: Diagnostic and Therapeutic Implications. Cancers. 2022;14(2):280. Available from: https://doi.org/10.3390/CANCERS14020280.

[2] Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37(4):471–484.

[3] Bell CC, Gilan O. Principles and mechanisms of non-genetic resistance in cancer. British Journal of Cancer. 2020;122:465–472.

[4] Pillai M, Hojel E, Jolly MK, Goyal Y. Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. Nature Computational Science 2023 3:4. 2023;3(4):301–313.

[5] Brown MS, Abdollahi B, Wilkins OM, Lu H, Chakraborty P, Ognjenovic NB, et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. Science Advances. 2022;8(31).

[6] Jain P, Bhatia S, Thompson EW, Jolly MK. Population Dynamics of Epithelial–Mesenchymal Heterogeneity in Cancer Cells. Biomolecules. 2022;12(3):348. Available from: https://doi.org/10.3390/biom12030348.

[7] Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. Nature Communications. 2019;10:5587. Available from: https://doi.org/10.1101/570341.

[8] Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer. 2021;3(3).

[9] Font-Clos F, Zapperi S, Porta CAM. Topography of epithelial–mesenchymal plasticity. Proceedings of the National Academy of Sciences. 2018;115(23):5902–5907.

[10] Hari K, Sabuwala B, Subramani BV, Porta CAM, Zapperi S, Font-Clos F, et al. Identifying inhibitors of epithelial–mesenchymal plasticity using a network topology-based approach. Npj Systems Biology and Applications 2020 6:1. 2020;6(1):1–12.

[11] Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. PLOS Computational Biology. 2015;11(11):1004569. Available from: https://doi.org/10.1371/journal.pcbi.1004569.

[12] Rashid M, Hari K, Thampi J, Santhosh NK, Jolly MK. Network topology metrics explaining enrichment of hybrid epithelial/mesenchymal phenotypes in metastasis. PLOS Computational Biology. 2022;18(11):1010687. Available from: https://doi.org/10.1371/JOURNAL.PCBI.1010687.

[13] Steinway SN, Zañudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. Npj Systems Biology and Applications. 2015;1:15014. Available from: https://doi.org/10.1038/npjsba.2015.14.

[14] George JT, Jolly MK, Xu S, Somarelli JA, Levine H. Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. Cancer Research. 2017;77(22):6415–6428.

[15] Ruscetti M, Dadashian EL, Guo W, Quach B, Mulholland DJ, Park JW, et al. HDAC inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. Oncogene. 2016;35(29):3781–3795.

[16] Celià-Terrassa T, Bastian C, Liu DD, Ell B, Aiello NM, Wei Y, et al. Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. Nature communications. 2018;9(1):5005.

[17] Subbalakshmi AR, Kundnani D, Biswas K, Ghosh A, Hanash SM, Tripathi SC, et al. NFATc acts as a non-canonical phenotypic stability factor for a hybrid epithelial/mesenchymal phenotype. Frontiers in oncology. 2020;10:553342.

[18] Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK. Landscape of epithelial mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. ELife. 2022;11.

[19] Boareto M, Jolly MK, Goldman A, Pietilä M, Mani SA, Sengupta S, et al. Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. Journal of the Royal Society Interface. 2016;13(118):20151106.

[20] Jolly MK, Boareto M, Debeb BG, Aceto N, Farach-Carson MC, Woodward WA, et al. Inflammatory breast cancer: A model for investigating cluster-based dissemination. Npj Breast Cancer. 2017;3(1):1–7.

[21] Neelakantan D, Zhou H, Oliphant MUJ, Zhang X, Simon LM, Henke DM, et al. EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. Nature Communications. 2017;8:15773. Available from: `https://doi.org/10.1038/ncomms15773`.

[22] Yamamoto M, Sakane K, Tominaga K, Gotoh N, Niwa T, Kikuchi Y, et al. Intratumoral bidirectional transitions between epithelial and mesenchymal cells in triple-negative breast cancer. Cancer Science. 2017;108(6):1210–1222.

[23] Hitomi M, Chumakova AP, Silver DJ, Knudsen AM, Pontius WD, Murphy S, et al. Asymmetric cell division promotes therapeutic resistance in glioblastoma stem cells. JCI Insight. 2021;6(3).

[24] Tripathi S, Chakraborty P, Levine H, Jolly MK. A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. PLoS Computational Biology. 2020;16(2):1–27.

[25] Munsky B, Fox Z, Neuert G. Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. Methods. 2015;85:12–21.

[26] Pally D, Goutham S, Bhat R. Extracellular matrix as a driver for intratumoral heterogeneity. Physical Biology. 2022;19(4):043001. Available from: `https://doi.org/10.1088/1478-3975/AC6EB0`.

[27] Lovisa S, LeBleu VS, Tampe B, Sugimoto H, Vadnagara K, Carstens JL, et al. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. Nature Medicine. 2015;21(9):998–1009.

[28] Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009;459(7245):428–432.

[29] Strasen J, Sarma U, Jentsch M, Bohn S, Sheng C, Horbelt D, et al. Cell-specific responses to the cytokine TGF β are determined by variability in protein levels. Molecular systems biology. 2018;14(1):e7733.

[30] Jain P, Corbo S, Mohammad K, Sahoo S, Ranganathan S, George JT, et al. Epigenetic memory acquired during long-term EMT induction governs the recovery to the epithelial state. Journal of the Royal Society Interface. 2023;20(198):20220627.

[31] Mantzaris NV. 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. Biophysical Journal. 2007;92(12):4271–4288.

[32] Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, Allgöwer F. Identification of models of heterogeneous cell populations from population snapshot data. BMC bioinformatics. 2011;12:1–15.

[33] Shu CC, Chatterjee A, Dunny G, Hu WS, Ramkrishna D. Bistability versus bimodal distributions in gene regulatory processes from population balance. PLoS Computational Biology. 2011;7(8).

[34] Spetsieris K, Zygourakis K, Mantzaris NV. A novel assay based on fluorescence microscopy and image processing for determining phenotypic distributions of rod-shaped bacteria. Biotechnology and Bioengineering. 2009;102(2):598–615.

[35] Hasenauer J, Schittler D, Allgöwer F. Analysis and simulation of division-and label-structured population models: a new tool to analyze proliferation assays. Bulletin of mathematical biology. 2012;74:2692–2732.

[36] Schittler D, Allgöwer F, De Boer RJ. A new model to simulate and analyze proliferating cell populations in BrdU labeling experiments. BMC systems biology. 2013;7:1–6.

[37] Bhatia S, Monkman J, Blick T, Pinto C, Waltham A, Nagaraj SH, et al. Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. J Clin Med. 2019;8(6):893.

[38] Chertock A. A practical guide to deterministic particle methods. In: Handbook of numerical analysis. vol. 18. Elsevier; 2017. p. 177–202.

[39] Alvarez FE, Guilberteau J. A particle method for non-local advection-selection-mutation equations. arXiv preprint arXiv:230414210. 2023.

[40] Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. Proceedings of the National Academy of Sciences. 2013;110(45):18144–18149.

[41] Jia W, Deshmukh A, Mani SA, Jolly MK, Levine H. A possible role for epigenetic feedback regulation in the dynamics of the epithelial-mesenchymal transition (EMT. Physical Biology. 2019;16(6):066004. Available from: `https://doi.org/10.1088/1478-3975/ab34df`.

[42] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. Nature. 2006;444(7119):643–646.

[43] Vega S, Morales AV, Ocaña OH, Valdés F, Fabregat I, Nieto MA. Snail blocks the cell cycle and confers resistance to cell death. Genes & development. 2004;18(10):1131–1143.

[44] Corre G, Stockholm D, Arnaud O, Kaneko G, Viñuelas J, Yamagata Y, et al. Stochastic fluctuations and distributed control of gene expression impact cellular memory. PLoS One. 2014;9(12):e115574.

[45] Nordick B, Yu PY, Liao G, Hong T. Nonmodular oscillator and switch based on RNA decay drive regeneration of multimodal gene expression. Nucleic Acids Research. 2022;50(7):3693–3708.

[46] 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 Systems. 2018;6(5):593–603.

[47] Mantzaris NV. Stochastic and deterministic simulations of heterogeneous cell population dynamics. Journal of theoretical biology. 2006;241(3):690–706.

[48] Degond P, Mas-Gallic S. The weighted particle method for convection-diffusion equations. I. The case of an isotropic viscosity. Mathematics of computation. 1989;53(188):485–507.

[49] Waldherr S. Estimation methods for heterogeneous cell population models in systems biology. Journal of The Royal Society Interface. 2018;15(147):20180530.

[50] Spetsieris K, Zygourakis K. Single-cell behavior and population heterogeneity: solving an inverse problem to compute the intrinsic physiological state functions. Journal of biotechnology. 2012;158(3):80–90.

[51] Mantzaris NV, Daoutidis P, Srienc F. Numerical solution of multi-variable cell population balance models. III. Finite element methods. Computers & Chemical Engineering. 2001;25(11-12):1463–1481.

[52] Mantzaris NV, Daoutidis P, Srienc F. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. Computers & Chemical Engineering. 2001;25(11-12):1411–1440.

[53] Dormand JR, Prince PJ. A family of embedded Runge-Kutta formulae. Journal of computational and applied mathematics. 1980;6(1):19–26.