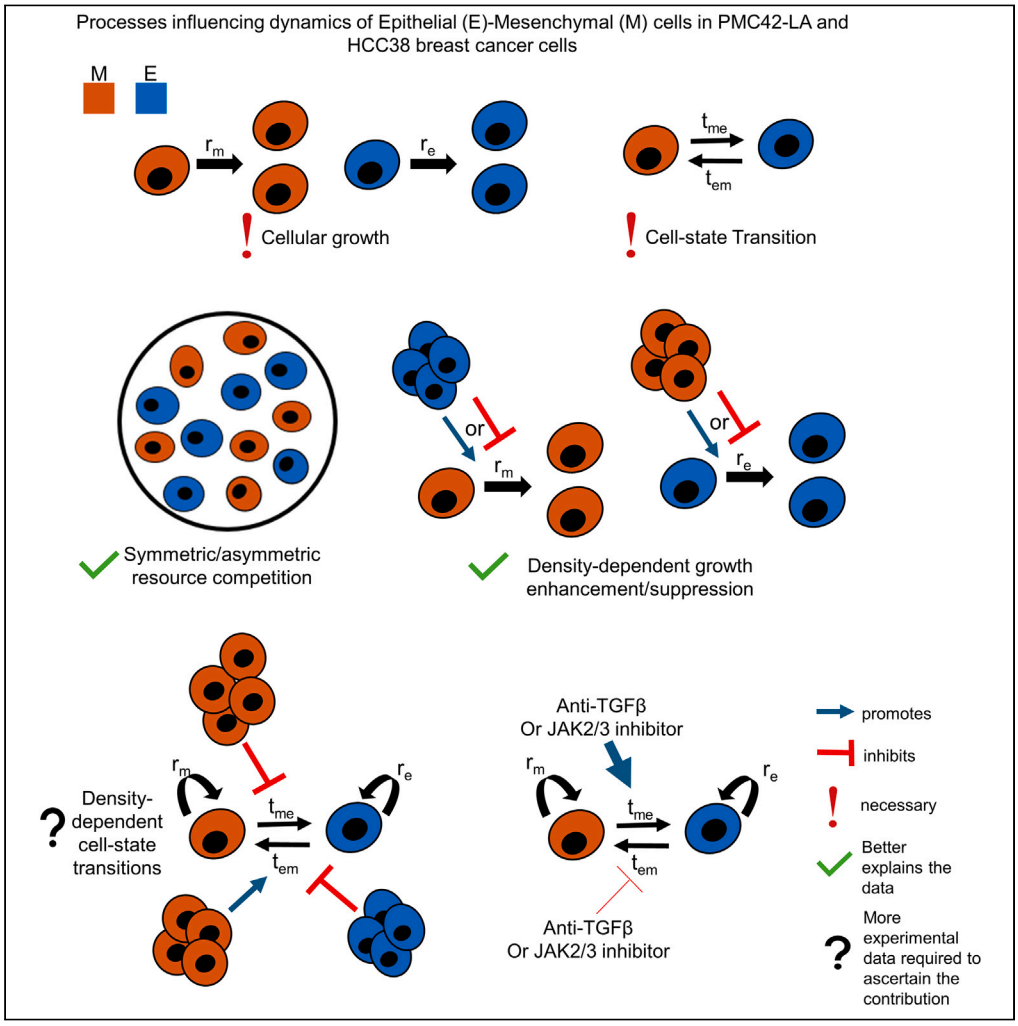
iScience

CellPress
OPEN ACCESS

Article

# Cell-state transitions and density-dependent interactions together explain the dynamics of spontaneous epithelial-mesenchymal heterogeneity




Paras Jain, Ramanarayanan Kizhuttil, Madhav B. Nair, Sugandha Bhatia, Erik W. Thompson, Jason T. George, Mohit Kumar Jolly

jason.george@tamu.edu (J.T.G.)
mkjolly@iisc.ac.in (M.K.J.)

## Highlights

Cell-state switching is essential to explain PMC42-LA, HCC38 cell population dynamics

Resource competition and density-dependent growth impact the E-M population dynamics

Cell-state switching rates altered by TGF-β, JAK2/3 inhibitor treatment in HCC38 cells

Jain et al., iScience 27, 110310
July 19, 2024 
https://doi.org/10.1016/j.isci.2024.110310

iScience

CellPress
OPEN ACCESS

Article

# Cell-state transitions and density-dependent interactions together explain the dynamics of spontaneous epithelial-mesenchymal heterogeneity

Paras Jain,[1,2] Ramanarayanan Kizhuttil,[3] Madhav B. Nair,[3] Sugandha Bhatia,[4] Erik W. Thompson,[5] Jason T. George,[2,6,*] and Mohit Kumar Jolly[1,7,*]

## SUMMARY

**Cancer cell populations comprise phenotypes distributed among the epithelial-mesenchymal (E-M) spectrum. However, it remains unclear which population-level processes give rise to the observed experimental distribution and dynamical changes in E-M heterogeneity, including (1) differential growth, (2) cell-state switching, and (3) population density-dependent growth or state-transition rates. Here, we analyze the necessity of these three processes in explaining the dynamics of E-M population distributions as observed in PMC42-LA and HCC38 breast cancer cells. We find that, while cell-state transition is necessary to reproduce experimental observations of dynamical changes in E-M fractions, including density-dependent growth interactions (cooperation or suppression) better explains the data. Further, our models predict that treatment of HCC38 cells with transforming growth factor β (TGF-β) signaling and Janus kinase 2/signal transducer and activator of transcription 3 (JAK2/3) inhibitors enhances the rate of mesenchymal-epithelial transition (MET) instead of lowering that of E-M transition (EMT). Overall, our study identifies the population-level processes shaping the dynamics of spontaneous E-M heterogeneity in breast cancer cells.**

## INTRODUCTION

Tumor heterogeneity still remains a major challenge in treating cancer. Population-level heterogeneity manifests as the occurrence of distinct cellular phenotypes resulting from genetic and non-genetic variability among single cells. While extensive genome-level characterization of many cancer types has identified driver mutations and directed the development of targeted drugs to contain tumor size, the emergence of resistance and tumor relapse is an ever-present threat. Drug-resistant populations emerge under evolutionary pressures that ultimately select for slowly proliferating pre-existent resistant cells or acquired resistance.[1] Recent empirical evidence has suggested that both pre-existent and acquired resistance can be found even in clonal populations with cells having identical genetic backgrounds.[2] Non-genetic heterogeneity and its consequent resistance therefore further complicate our understanding given the diversity of levels (proteome, metabolome) and cellular processes (epithelial [E]-mesenchymal [M] transition, oxidative phosphorylation/glycolysis, stem cell/non-stem cell) at which cells can manifest epigenetic (non-genetic) differences.[3]

Switching among E and M states represents one canonical example of non-genetic cellular heterogeneity and has been reported in embryonic development, wound healing, and diseased states, including fibrosis and cancer metastasis. The switches among E-M states have been shown to occur by mutual interaction between several regulatory players and signaling pathways.[4,5] Further, single-cell characterization at the transcriptome and chromatin levels has revealed the extent of E-M transitions (EMTs) and M-E transitions (METs) in response to different concentrations of growth factors and the duration of their exposure.[6–8] Apart from molecular characterization of EMT and MET, few population-level studies evaluated E-M state composition of cancer cells and their corresponding changes over time. The E and M cells are found to be distributed in variable fractions across distinct cell lines, even within the same cancer type. For example, luminal-like and claudin-low breast cancer cell lines were shown to be composed of a majority of epithelial and mesenchymal cells, while basal-like cell lines were enriched in cells co-expressing E and M markers.[9] The phenotypic composition of E-M cells can also change in experiments tracking clonal populations of single cells derived from a parental cell line.[10] Similarly, the PMC42-LA breast cancer cell line exhibits a stable composition (80:20 ratio of

[1]Department of Bioengineering, Indian Institute of Science, Bangalore, India
[2]Department of Biomedical Engineering, Texas A&M University, College Station, TX, USA
[3]Indian Institute of Science Education and Research, Kolkata, India
[4]School of Biomedical Science, Queensland University of Technology (QUT) at Translational Research Institute, Woolloongabba QLD 4102, Australia
[5]Diamantina Institute, The University of Queensland, Brisbane QLD, Australia
[6]Center for Theoretical Biological Physics, Rice University, Houston, TX, USA
[7]Lead contact
*Correspondence: jason.george@tamu.edu (J.T.G.), mkjolly@iisc.ac.in (M.K.J.)
https://doi.org/10.1016/j.isci.2024.110310

E:M cells); however, its parental cell line PMC42-ET is M cell predominant.[11,12] Consequently, this variability necessitates a more detailed framework capable of explaining such differences, yet the mechanisms determining phenotypic composition at a population level and its dynamical evolution currently remain unknown.

Experimental evidence suggests temporally varying E-M cell fractions when E-M populations are either isolated or mixed in different proportions.[9,12–14] E, M, and hybrid E/M subpopulations (using canonical E-Cadherin and Vimentin levels) were simultaneously observed in castration-resistant metastatic prostate cancer cell lines.[13] Inhibition of HMGA2 destabilized M phenotypic state and reduced E-M transition thereby lowering down the proportion of M cells in the population. Similarly, in HCC38 breast cancer cells, the phenotypic composition of E-M cells in the population was found to influence the extent of EMT and/or MET. Further, the inhibition of the transforming growth factor β (TGF-β) signaling pathway and blocking of JAK2/3 diminished the effects that the M cells had on EMT. Thus, the authors suggested that cell-cell communication was at least partly responsible for influencing cell-state transitions at the population level. Unfortunately, however, the aforementioned studies did not explicitly report the growth rates of E-M cell populations in each scenario, thus missing out inadvertently on contributions of different growth rates on E-M heterogeneity dynamics. Measuring any change(s) in growth rates is crucial for two reasons: (1) growth rate differences among subpopulations impact the population-level composition[15,16] and (2) EMT is also known to slow down cell proliferation,[17,18] and, thus, inhibition of EMT may not only reduce cell-state transition but may increase the proliferation rate.[19] Therefore, the relative contribution of various interconnected cell-autonomous and non-cell-autonomous effects on population dynamics of EMT and MET remains poorly understood.

Here, we investigated which population-level processes are necessary to explain the experimentally observed time course data on spontaneous emergence of E-M heterogeneity in PMC42-LA and HCC38 breast cancer lines including (1) differential growth rates, (2) cell-state transitions,[13,20–22] and (3) cell-cell communication via modulating either growth or cell-state-transition rates.[9,23–28] We compared a number of distinct mathematical models combining one or more of the above cellular processes according to their goodness of fit to the experimental data, while optimizing model parameterization restricted to biologically relevant ranges. Applying the information theoretic-based corrected Akaike information criteria (AICc) to select the best-fit model, we found that including E-M cell-state transitions was necessary, albeit at slow rates, to explain the emergence of dynamic E-M heterogeneity observed in both PMC42-LA and HCC38 cells. Further, on incorporating cell-state transitions, the models having population density-dependent growth rates were better in explaining the data than the models with population density-dependent cell-state-transition rates. The diminished role of cellular interactions influencing cell-state transitions was substantiated by fitting TGF-β and JAK2/3 inhibitor treatment data as well. We performed uncertainty analysis of models in explaining the data by commenting on their false-positive selection due to noisy experimental data and sensitivities to the initial proportion of E and M cells in the experiments. Lastly, we devised a selection criterion to identify the next most informative time points for which future experimental data will optimally improve the identifiability (95% confidence bounds) of the estimated best-fit model parameters. Overall, our results demonstrate how phenotypic heterogeneity in breast cancer cells emerges as a consequence of cell-state transitions together with the influence of heterogeneous subpopulations on each subpopulation's growth rate.

## RESULTS

### Cell-state transitions are necessary for explaining spontaneous EMP in PMC42-LA breast cancer cells

Recent experiments reported that PMC42-LA breast cancer cells exhibited an 80:20 ratio of E and M subpopulations based on epithelial cell adhesion molecule (EpCAM) levels ($EpCAM^{high}$ and $EpCAM^{low}$, respectively).[12] When both subpopulations were isolated and cultured independently, they recapitulated the original parental distribution of an 80:20 E:M ratio after eight weeks of time (culture 1 and 2 in Figure 1A – experimental data). These observations of spontaneous E-M plasticity (EMP) motivated us to investigate the role of distinct population-level processes (differential growth, cell-state transition, cell-cell communication) in governing these dynamics. Thus, we designed a set of mathematical models, each including different combinations of these processes, and compared their fit with the experimental data (Table 1). The model fitting was performed for each replicate simultaneously (please refer to the STAR Methods for procedure on model fitting and parameter optimization).

We observed that only a difference in the net growth rates (the difference of division and death rates) of E ($r_e$) and M ($r_m$) cells could not reproduce the dynamics and eventual 80:20 ratio of $EpCAM^{high}$ and $EpCAM^{low}$ subpopulations when the initial population consists of either M or E cells (model G with E net growth rate [$r_e$] ≥ M net growth rate [$r_m$]). This inaccuracy led us to further consider a scenario that also accounts for the proliferation rates being dependent on non-cell-autonomous effects. Thus, we considered population density-dependent cooperative or suppressive effect of one subpopulation on growth of the other, besides intrinsic growth differences between E and M subpopulations (Table 2 presents the currently known mechanisms of cell-nonautonomous growth regulations and associated references). Two additional parameters were introduced: σ (the influence of E cells on growth rate of M cells) and μ (the influence of M cells on growth rate of E cells) parameters. However, this growth influence (GI) model also failed to adequately explain the observed data (Figure 1A). In particular, we saw a sudden change in the fraction of M ($EpCAM^{low}$) cells for the GI model around the second week, an observation not seen in experimental data. This sudden change could be attributed to the threshold fraction of E and M populations required for influencing the growth rate of the other subpopulation. Together, these simulations show that a difference in intrinsic growth rate—with or without the influence of one subpopulation on growth rate of the other—alone cannot explain observed experimental data in PMC42-LA breast cancer cells.

We next hypothesized that including cell-state transitions between E and M phenotypes can help explain the experimental data more accurately. Thus, we added cell-state transition terms (E-to-M and M-to-E transitions) to the aforementioned models, resulting in growth and transition (G&T) and growth influence and transition (GI&T) models with $t_{em}$ and $t_{me}$ as the E-to-M and M-E per-cell transition rates, respectively (Table 1). Both G&T and GI&T gave significantly improved model fits to the experimental data (Figure 1A). However, these

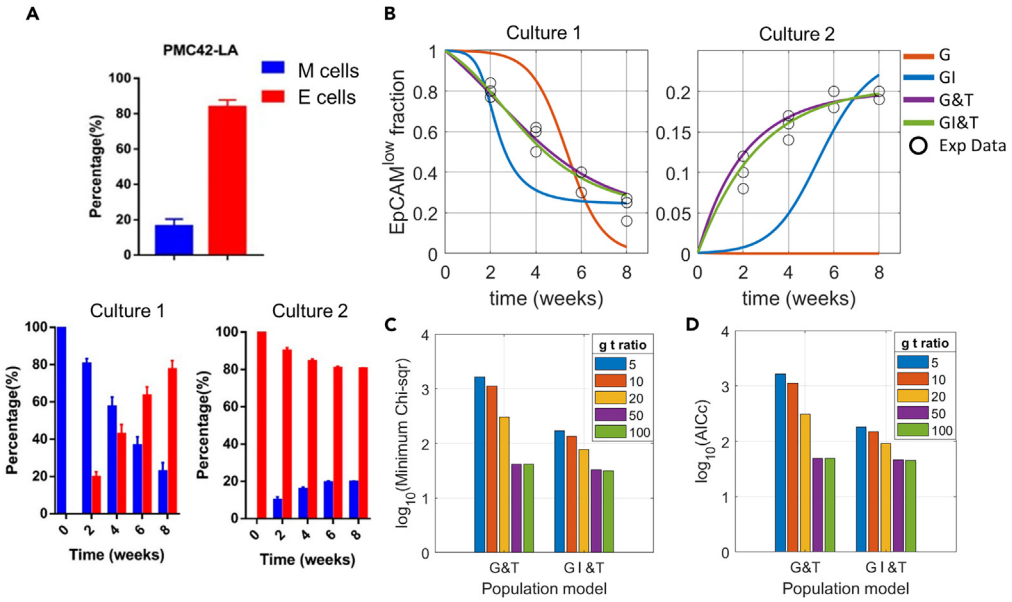



**Figure 1. Necessity of cell-state transition in explaining spontaneous EMP in PMC42-LA breast cancer cells**

(A) Distribution of EpCAM$^{high}$ (epithelial) and EpCAM$^{low}$ (mesenchymal) cells in PMC42-LA cell; changing phenotypic distribution of the population in cell cultures of isolated E and M cells from the parental cell line. 100% M and 100% E cells at week 0 in culture 1 and 2, respectively (Figure adopted from Bhatia et al.[12]).

(B) Fits to the experimental data using population models that consider – (1) Growth of E-M subpopulations (G – growth), (2) E-M subpopulations mutual influence on each other's growth (GI – growth influence), (3) Growth and cell-state transition among E-M subpopulations (G&T – growth and transition), and (4) E-M subpopulations influence on each other's growth with cell-state transitions (GI&T – growth competition and transition). The subpopulations mutual influence can be either cooperative or suppressive. Please refer to Table 1 for model formalism.

(C) Improvement in goodness of fit (measured by chi-square values) of G&T and GI&T models when cell-state-transition rates are allowed to take smaller values with respect to mesenchymal cell growth rates (lower bound on normalized transition rates = 1/(g t ratio); g t ratio: growth-to-transition ratio).

(D) Model selection using Corrected Akaike Information Criteria (AICc) values. As with minimum chi-square values (B), GI&T model has lower AICc values than G&T for a given value of g t ratio. At g t ratio of 50, the GI&T AICc is 46.38 while G&T AICc is 49.19. g t ratio = 50 for model fits in (A). Mesenchymal cell growth rate has been used to non-dimensionalize the model equation and its value is set to (1/50) hrs$^{-1}$ to rescale the time units of model output and compare temporal dynamics with experimental data.

improvements in model fitting were observed only when the E-M and M-E transition occurred at rates much slower than growth rates of mesenchymal cells (Figures 1B and S1). Here, the parameter "g-t ratio" (growth-to-transition rate ratio) determined the lower bound on the search range for E-M and M-E transition rates during the parameter-optimization process. The larger the g-t ratio, the smaller the lower bound on transition rates (Table 3). Thus, although cell-state transitions are necessary to explain the experimental observation, our findings suggest that they must occur at slower rates relative to cell division to explain the experimentally observed data.

**Table 1. Model proposed for fitting spontaneous E-M plasticity seen in PMC42-LA cells**

| Model | Description |
|---|---|
| G&T | Growth and transition |
| GI&T | Growth influence and transition |
| G&T-Mr | Growth, transition, and M cells retention (M cells reducing M-E transition) |
| G&T-Er | Growth, transition, and E cells retention (E cells reducing E-M transition) |
| G&T-EMr | Growth, transition, E and M cells retention |
| G&T-Mi | Growth, transition, M cells' influence on E cells for E-M transition |
| G&T-Mi-Mr | Growth, transition, M cells' influence on E cells for E-M transition, and M cells retention (M cells reducing M-E transition) |

**Table 2. Literature evidence for the cell-non-autonomous interactions among E-M cells**

| Regulation | Mechanism | Reference |
|---|---|---|
| Growth influence (cooperation/suppression) | paracrine signaling of metabolites—lactate, β-hydroxybutyrate, estradiol | Farrokhian et al.[24], Emond et al.[27], and Noble et al.[28] |
| Resource competition | Growth dependence on pH and glucose levels of the culture media | Emond et al.[27], Freischel et al.[29] |
| Epithelial retention | BMP signaling in E cells | Scheel et al.[30] |
| Mesenchymal retention | Autocrine TGF-β, JAK2/STAT3, canonical and non-canonical Wnt signaling in M cells | Gregory et al.[26], Scheel et al.[30], and Katsuno et al.[31] |
| Mesenchymal influence | Canonical (SIX1-SHH-Ptch/Smo-GLI activation) or non-canonical (SIX1-GLI activation via TGF-β) | Neelakantan et al.[25], Hapach et al.[32] |

In the earlier discussion, one could argue that since the experiment started with pure population of EpCAM$^{low/high}$ cells, the necessity of cell-state transitions between EpCAM$^{low}$ and EpCAM$^{high}$ states is trivial to bring in cellular heterogeneity seen in the data. However, as the experimental cell sorting of a subpopulation from the parental cell line contains some impurity of other subpopulations, we also considered that the initial population which is stated as 100% EpCAM$^{high}$ has 99.9% of EpCAM$^{high}$ (E) and 0.1% of EpCAM$^{low}$ (M) cells and vice-versa for the 100% EpCAM$^{low}$ case. But, even after having impure/heterogeneous population to start with, the models capturing just growth dynamics, and not cell-state transition, cannot explain the PMC42-LA cells data.

Since parameter values within their 95% confidence could give equally good fit to the experimental data, we next looked at how model parameters are co-variable with each other. Perhaps intuitively, in order for the G&T model to explain a significant long-run (20%) contribution of M cells in the population, an increase in E growth rates ($r_e$) must lead to a concomitant increase in E-M ($t_{em}$) and decrease in M-E ($t_{me}$) transition rates (Figure S2). This interdependence between growth and transition rates was subsequently weakened through the inclusion of additional growth-influencing parameters (σ and μ) in the case of the GI&T model (Figure S3). In the GI&T model, σ (the influence of E cells on the growth rate of M cells) became more negative with increasing $r_e$ values, and thus E cells could increasingly support M cells growth when E cells themselves divided faster. Similarly, the suppression of E cell growth by M cells increased when E cells divided faster (Figure S3). Effectively, this addition allows intrinsic growth rates and GI parameters to offset one another's effects and as a result contributed to unidentifiability of parameters. GI (σ and μ) parameters barely influenced transition rates ($t_{me}$ and $t_{em}$) in most of their 95% confidence range.

We next calculated AICc values to perform model selection that achieves a trade-off between the goodness of fit and model complexity (number of parameters). We found that in the g-t ratio regime where both the G&T and GI&T models performed their best, GI&T exhibited slightly improved AICc values (Figure 1C). These findings suggested that the most probable process that gave rise to the observed experimental phenomenon included (1) intrinsic growth rate differences between E and M cells, (2) cooperative/suppressive growth interaction, and (3) cell-state transitions.

## Cell-state transition, together with population density-dependent GI of subpopulations, explains spontaneous EMP in HCC38 cells

We next focused on another set of EMT population dynamics experiments where co-cultures of different starting ratios of E and M cells were tracked, and the temporal dynamics of phenotypic heterogeneity was recorded over time[9]. In these experiments, the authors co-cultured various percentage of Venus (dye)-labeled EpCAM$^{high}$ cells with Venus-non-labeled EpCAM$^{low}$ cells, and they tracked increasing fractions of Venus-labeled EpCAM$^{low}$ cells in the Venus-labeled population over 12 days (culture 1 to 3 in Figure 2A). Similar co-cultures were

**Table 3. Parameter search ranges while optimizing models proposed for fitting PMC42-LA E-M heterogeneity emergence data**

| Parameter | Description | Search range |
|---|---|---|
| $r_m$ | M cell growth rate (1/h) | $r_m$ |
| $r_e$ | E cell growth rate (1/h) | $[r_m \quad 3r_m]$ |
| $t_{me}$ | M-E transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio} \quad r_m]$ |
| $t_{em}$ | E-M transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio} \quad r_m]$ |
| $\sigma$ | E cells influence on M cells growth rate | [−1 1] |
| $\mu$ | M cells influence on E cells growth rate | [−1 1] |
| $\gamma$ | strength of E cells to reduce E-M transition | [0 1] |
| $\delta$ | strength of M cells to reduce M-E transition | [0 1] |
| $\theta$ | strength of M cells to enhance E-M transition | [1 4] |

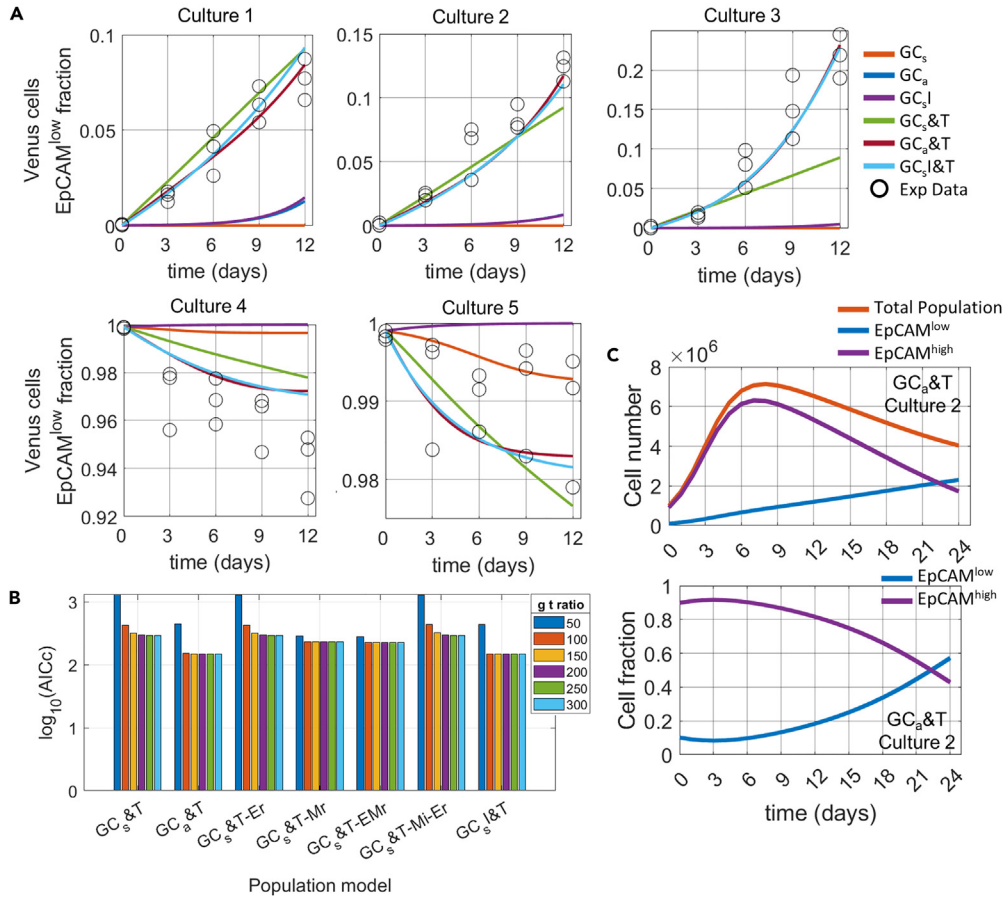



**Figure 2. Necessity of both mutual growth competition/influence and cell-state transitions among subpopulations to explain spontaneous EMP in HCC38 cells**

(A) Model fits to experimental data of different culture conditions depending upon the initial mixture of Venus label and EpCAM$^{high/low}$ status. The models here capture the following mechanisms/interactions among E and M subpopulations – (1) Symmetric resource competition for growth ($GC_s$), (2) Asymmetric resource competition for growth ($GC_a$), (3) Symmetric resource competition and growth influence ($GC_sI$). We then added cell-state transitions to the above models to obtain $GC_s$&T, $GC_a$&T, and $GC_sI$&T models. Please refer to methods and Table 4 for model formalism (The models considered here have logistic growth term unlike those used to fit data in Figure 1, which considered exponential growth).

(B) Comparison of corrected Akaike information criteria (AICc) values across expanded set of models that also include models that consider mutual influence of subpopulations through changes in the transition rates as the population frequency distribution changes with time. Here, "Er" denote that the epithelial cells lower down their E-to-M transition rate with their increasing frequency in the population. Similarly, such retention effect is considered for the M cells as well ($GC_s$&T-Mr). Model $GC_s$&T-Mi-Er incorporates retention effect of both E and M populations. Term "Mi" in model G&T-Mi-Er indicates that M cells increase the E-to-M transition rates with their increasing frequency in the population.

(C) Temporal dynamics of EpCAM$^{low}$ and EpCAM$^{high}$ cell number and fraction in the population from model $GC_a$&T using best-fit parameters.

performed for varied percentage of Venus-labeled EpCAM$^{low}$ cells with Venus-non-labeled EpCAM$^{high}$ cells to track the increasing fraction of EpCAM$^{high}$ cells (decreasing fraction of EpCAM$^{low}$ cells) in the Venus-labeled population over 12 days (culture 4 and 5 in Figure 2A). The experiments demonstrated that the measured fraction of E cells that transitioned to the M state was higher at each measured time point when there was a higher initial frequency of M cells in the co-culture population. Similarly, the measured fraction of M cells that transitioned to E cells was higher when there was a higher initial frequency of E cells in the co-culture population. Relatively speaking, the impact of increased percentage of E cells on MET was less than that of increased percentage of M cells on EMT. These experimental data support the idea that interactions between E and M populations can modulate their cell-state-transition rates.

To explain this experimental data, we asked our original question: is it necessary to incorporate cell-state transition in adequately describing the experimental data? The models considered here are analogous to those considered for fitting PMC42-LA cell data; however, we included logistic (instead of exponential) growth term in all the models. We made this change to mimic experimental conditions where the cells were continuously grown without passaging. We considered two types of logistic growth: (1) symmetric competition—where E and M cells consume equal number of resources—and (2) (two-species ecological) asymmetric competition—where E and M cells consume unequal number of resources. The experimental study reported the doubling times of E and M cells to be 35 and 54 h, respectively, and therefore the growth rate ($r_e$ and $r_m$)

**Table 4. Model proposed for fitting spontaneous E-M plasticity seen in HCC38 cells**

| Model | Description |
|---|---|
| $GC_s$&T | Logistic growth (symmetric competition) and transition |
| $GC_a$&T | Logistic growth (asymmetric competition) and transition |
| $GC_s$&T-Er | Logistic growth (symmetric competition), transition, and E cells retention (E cells reducing E-M transition) |
| $GC_s$&T-Mr | Logistic growth (symmetric competition), transition, and M cells retention (M cells reducing M-E transition) |
| $GC_s$&T-EMr | Logistic growth (symmetric competition), transition, and both E and M cells retention |
| $GC_s$&T-Mi-Er | Logistic growth (symmetric competition), and M cells' influence on E cells for E-M transition, and M cells retention |
| $GC_s$I | Logistic growth (symmetric competition) and growth influence of E and M subpopulations on each other |
| $GC_s$I&T | Logistic growth (symmetric competition), growth influence of subpopulations, and cell-state transitions |
| $GC_s$&T3 | Three states (E, M, and H) logistic growth (symmetric competition) and cell-state transitions |
| $GC_s$&T3-L | Three states (E, M, and H) logistic growth (symmetric competition) and cell-state transitions (no direct transition between E and M states) |

parameters for all models were set constant as per the aforementioned values and were not included in the model-fitting exercise. The HCC38 cells data also presented the impurity of Venus-labeled EpCAM$^{low}$ (EpCAM$^{high}$) cells in the sorted Venus-labeled EpCAM$^{high}$ (EpCAM$^{low}$) subpopulations at day 0 due to inefficient cell sorting,[9] which we accounted for in the initial population distribution in our model-fitting exercise.

We first considered models that captured only population dynamics due to growth differences, competition, and GI (cooperative/suppressive): (1) $GC_s$ (logistic growth symmetric competition), (2) $GC_a$ (logistic growth asymmetric competition), and (3) $GC_s$I (logistic growth symmetric competition with subpopulations influence on growth) (Table 2 presents the currently known mechanisms of cell-nonautonomous growth regulations and their associated references). However, these models could not explain the experimental data well (Figure 2A; please refer to the STAR Methods for procedure on model fitting and parameter optimization). The inability of these models to explain the data remained even after we made either $r_m$ or both $r_m$ and $r_e$ non-constants and optimized for them along with other parameters during the fitting exercise (Figure S4). Inclusion of cell-state transitions led to significant improvement in the goodness-of-fit (lower chi-square) values, although again at large g-t ratio (Figures S5A and S5B; model $GC_s$&T, $GC_a$&T, and $GC_s$I&T). We again highlight that the necessity of cell-state transitions is nontrivial to generate cellular heterogeneity and, therefore, explain the data as the cellular population was already heterogeneous (impure) at day 0.

To decipher the possible interactions among E-M populations, we expanded our models set to include subpopulation interactions that modulate transition rates apart from GI (please refer to Table 4 and the STAR Methods for detailed model descriptions). We considered two effects, retention and influence, in the ability of E and M subpopulations to influence cell-state-transition rates. By *retention* (denoted by the lower case "r") we mean that a subpopulation reduces its cell-state-transition rate to another subpopulation (phenotype) when in majority. On the contrary, the *influence* effect (denoted by the lower case "i") of E/M subpopulations enhances cell-state-transition rates of another subpopulation (phenotype) to its own when in majority. The models considered include retention in either E or M cells or both ($GC_s$&T-Er, $GC_s$&T-Mr, and $GC_s$&T-EMr) and influence of M cells on E cells ($GC_s$&T-Mi-Er). These specific cellular interactions affecting cell-state transitions were considered along the lines of existing reports (Table 2 presents the currently known mechanisms of cell-nonautonomous transition regulations and their associated references).

The proposed models were checked for parameter identifiability using *a priori* and *a posteriori* methods.[33–35] *A priori* structural identifiability of models using a differential algebra approach resulted in all models being structurally unidentifiable in the global parametric ranges. However, *a posteriori* identifiability analysis using the profile likelihood method resulted in all parameters across models to exhibit

**Table 5. Parameter search ranges while optimizing models proposed for fitting HCC38 E-M heterogeneity emergence data**

| Parameter | Description | Search range |
|---|---|---|
| $r_m$ | M cell growth rate (1/h) | $r_m$ |
| $r_e$ | E cell growth rate (1/h) | $r_e$ |
| $t_{me}$ | M-E transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $t_{em}$ | E-M transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $t_{mh}$ | M-H transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $t_{hm}$ | H-M transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $t_{eh}$ | E-H transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $t_{he}$ | H-E transition rate (1/h) | $[\frac{r_m}{g\ t\ ratio}, r_m]$ |
| $\alpha$ | E cells influence on M cells growth rate | [0,10] |
| $\beta$ | M cells influence on E cells growth rate | [0,10] |
| $\gamma$ | strength of E cells to reduce E-M transition | [0,1] |
| $\delta$ | strength of M cells to reduce M-E transition | [0,1] |
| $\theta$ | strength of M cells to enhance E-M transition | [1,4] |
| $\sigma$ | E cells influence on M cells growth rate | [-1,1] |
| $\mu$ | M cells influence on E cells growth rate | [-1,1] |
| K | Common carrying capacity (cells) | $[10^6, 20x10^6]$ |
| $K_e$ | E cells carrying capacity (cells) | $[10^5, 20x10^6]$ |
| $K_m$ | M cells carrying capacity (cells) | $[10^5, 20x10^6]$ |

well-defined upper and/or lower bounds (Figures S4E and S4F). Thus, the data along with the well-defined biologically relevant parameter range (Table 5) were together able to resolve parameters in well-defined intervals.

Incorporation of population-specific effects on transition rates did not change the dependence of models incorporating cell-state transitions to require large g-t ratio to adequately fit the data (Figure S5B). All the models predict the E-M transition to be much faster than the M-E transition (transition rates $t_{em} \gg t_{me}$), which was also reported in the experimental study (Figure S5A).

On performing model selection, the use of AICc resulted in model $GC_a$&T (which considered asymmetric growth competition with basal cell-state transitions) and $GC_s$I&T (which considered symmetric growth competition with influence on growth by subpopulations and basal cell-state transitions) as being the best models with little difference of AICc values among themselves (Figure 2B). Other independent check of goodness of fit of $GC_a$&T and $GC_s$I&T was done by analyzing (1) number of data points normalized chi-square values and (2) *p* value of the chi-square statistic (Figures S5C and S5D). While a normalized chi-square value closer to 1 indicates good fit and least overfitting, *p* value of the chi-square statistic tells about the probability that the chi-square should exceed a particular value by chance (refer to the STAR Methods for more details on these two metrics). Both $GC_a$&T and $GC_s$I&T models had closest-to-1 normalized chi-square values and the greatest *p* value of chi-square statistic which indicate their greater ability to explain the data. Making either $r_m$ or both $r_e$ and $r_m$ as non-constant(s) and free parameter(s) still rendered $GC_a$&T and $GC_s$I&T as the best-fit models (Figure S6).

Focusing specifically on the parameter co-variability of the $GC_a$&T model 95% confidence range analysis using profile likelihood, we observed that growth competition affected the carrying capacity of E and M cells (Figure S7). With increasing growth competition to M cells by E cells ($\alpha$), $K_e$ was increased and $K_m$ was reduced. The reduced proliferation capacity ($K_e$) of E cells was substantiated by more M-E transitions, and less E-M transition with reduced growth competition to E cells by M cells ($\beta$) (Figure S7). Inversely, increasing growth competition to E cells by M cells ($\beta$) led to enhanced proliferation capacity of E cells ($K_e$) with concomitant increased E-M transitions, and diminished M-E transitions with reduced growth competition to M cells by E cells ($\alpha$) (Figure S7). Similarly, for the $GC_s$I&T model, higher growth-suppressive effect of M cells on E cells ($\mu$) led to increasing carrying capacity (K), and M-E transition rates; and larger cooperative interaction by E cells ($\sigma$) led to reduced E-M transition rates and increased M-E transition rates (Figure S8). However, the relations among parameters are based on their co-variability following a best fit to the data, and so it is therefore intractable to infer causal relations (independent/dependent) among parameters solely from the model-fitting exercise.

We next looked at the temporal dynamics of overall $EpCAM^{low}$ and $EpCAM^{high}$ cells, irrespective of their Venus label, for the best-fit model $GC_a$&T. $EpCAM^{high}$ cell numbers that had grown to large number in the first 10 days of culture started to drop down because of increasing resource competition from the mesenchymal population (Figure 2C top panel). These negative growth trends for E cells and positive trends for M cells led to an increasing long-run population fraction of M cells (Figure 2C bottom panel). Similarly, even for the $GC_s$I&T model, where E

**Table 6. Model-wise parameter list (corresponding to models proposed for fitting PMC42-LA E-M heterogeneity emergence data)**

| Parameter index | G&T | GI&T | G&T-Er | G&T-Mr | G&T-EMr | G&T-Mi-Mr |
|---|---|---|---|---|---|---|
| 1 | $r_e$ | $r_e$ | $r_e$ | $r_e$ | $r_e$ | $r_e$ |
| 2 | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ |
| 3 | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ |
| 4 | | $\sigma$ | $\gamma$ | $\delta$ | $\gamma$ | $\delta$ |
| 5 | | $\mu$ | | | $\delta$ | $\theta$ |

cells cooperate/enhance M cells growth and M cells supress E cells growth, M cells dominate the population frequency in the long run (Figure S9). However, the parental HCC38 cells had an EpCAM$^{high}$:EpCAM$^{low}$ ratio of 90:10. Therefore, we argue that such dominance of EpCAM$^{high}$ cells in the cell line can be reasonably explained by keeping E cells to be in the exponential phase growth, which provides them with an additional advantage over the large growth-suppressive influence of M cells and high E-M transition rates.

Although the models GC$_a$&T and GC$_s$I&T fitted the experimental data to a good extent, they were not able to completely follow the cell-fraction trajectories of culture 4 condition, even after making $r_e$ and $r_m$ as free parameters (Figures 2A and S10). This prompted us to consider simultaneous influence of density-dependent growth and transitions in population dynamics. Therefore, we added "epithelial retention," "mesenchymal retention," and "mesenchymal influence" density-dependent effects on transition rates to (asymmetric) growth competition (GC$_a$&T) and GI (GC$_s$&T) models. We found that model fits to the culture 4 data improved when M retention effects were coupled with either growth competition or influence models (Figure S11). Therefore, HCC38 cells seems to possess both density-dependent growth and cell-state transition effects, particularly M cell retention. However, the transition rate ($t_{me}$) depends weakly on density as even without density-dependence of transition (and just with density-dependent growth) the culture 4 data are well explained.

To test whether the impact of E and M subpopulation frequencies on the cell-state-transition rates improved the model fit to experimental data for PMC42-LA cells, we expanded the list of models to include additional terms corresponding to the effects of phenotype on their retention (i.e., E cells prevent their transition to M) and influence on others (i.e., E cells promote MET in M cells), respectively, and fitted them to experimental data (refer Table 1 and the STAR Methods for model formalism). Parameter identifiability analysis using analytical differential algebra method results in all models to be globally identifiable conditional on knowing $r_m$ and/or $r_e$ (SI Text). Consequently, parameter identifiability analysis using an *a posterior* profile likelihood method showed that given the experimental data and relevant bounds on the parameters (Table 3), all model parameters had either well-defined upper or lower bounds, with the exception of the models G&T-Mi and G&T-Mi-Mr, whose transition influence parameters did not have both upper and lower bounds (Figures S12B and S12C and Table 6). In the model set, the model G&T-EMr gave a significantly good fit at smaller g-t ratio values (Figure S12A). Since the g-t ratio determines the lower bound on basal transition rates ($t_{me}$ and $t_{em}$), smaller g-t ratio values well-describing the data are understandable given that the added retention effect acts to reduce the effective transition rate ($t_{me}(1-\delta)$ and $t_{em}(1-\gamma)$). This allows the G&T-EMr model to attain larger estimates of $t_{me}$ and $t_{em}$ parameters while fitting the data. For larger g-t ratio values, we found that G&T and GI&T had lower chi-square than other models, with GI&T having the lowest AICc, close-to-1 normalized chi-square and *p* value of chi-square statistic close to zero.(Figure S12A). These trends further strengthen that the models incorporating either independent or mutually dependent subpopulation growth rates, together with basal cell-state-transition rates, are best to explain the EMP dynamics seen in PMC42-LA cells.

We limited our analysis to two cell subpopulations/states only as the experimental data did not include hybrid states. However, mathematical analysis can detect the presence and role of hybrid states in the overall E-M population dynamics. Therefore, we considered three-states versions of G&T and GC$_s$&T models to fit the PMC42-LA and HCC38 cells data, respectively. Particularly, we considered four different forms of three-states models depending on whether (1) the transition between E and M cell states is direct or only possible through immediate hybrid E/M state (T3 and T3-L, here "L" stands for linear) and (2) the hybrid cell fraction is accounted in the experimental data either in E or M cells fraction (EH or HM) (Figure S13A) (please refer to the STAR Methods for model description). We found that, for HCC38 cells data, the three-state models GC$_s$&T3-EH and GC$_s$&T3-L-EH that include hybrid cells concomitantly present within E cell fraction fit the experimental data better than the two-state GC$_s$&T model (Figure S13C). Nonetheless, the three-state GC$_s$&T models were not able to give a better fit than two-states density-dependent growth models (GC$_a$&T and GC$_s$I&T) (compare AICc and *p* values in Figures 2B and S5D with Figure S13C). Furthermore, since the data did not explicitly capture the proportion and dynamics of the hybrid state, the parameters of the three-states GC$_s$&T model (GC$_s$&T3-L-EM) were poorly constrained in comparison to the other two-states model in our set (Figures S5F and S14). Similarly, even though three-state G&T model (G&T3-L-EH) fitted the PMC42-LA data better than two-state G&T model (G&T2) and gave a comparable fit as the density-dependent growth model GI&T (Figure S13B), the parameters of G&T3-L-EH model were again poorly constrained when compared to other two-states models (Figure S14). Therefore, incorporation of the hybrid cell state in the model does improve model fit, yet more complex interaction (density-dependent growth) is required to better explain the data.

## Changes in basal cell-state-transition rates can explain population dynamics of TGF-β and JAK2/3 inhibitor treatment data in HCC38 cells

To investigate whether cells communicate with each other through paracrine signaling and influence E-M and M-E transition dynamics, Yamamoto et al. further used varied EMT inhibitors to inactivate the crosstalk between EpCAM$^{low}$ and EpCAM$^{high}$ cells.[9] They observed that the

**Table 7. Model-wise parameter list corresponding to models proposed for fitting HCC38 E-M heterogeneity emergence data**

| Parameter index | $GC_s$&T | $GC_a$&T | G&T-Er | G&T-Mr | G&T-EMr | G&T-Mi-Er | $GC_s$I | $GC_s$I&T | $GC_s$&T3 | $GC_s$&T3-L |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{me}$ | $t_{mh}$ | $t_{mh}$ |
| 2 | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{em}$ | $t_{hm}$ | $t_{hm}$ |
| 3 | K | $\alpha$ | K | K | K | K | $\sigma$ | K | $t_{he}$ | $t_{he}$ |
| 4 | | $\beta$ | $\gamma$ | $\delta$ | $\delta$ | $\gamma$ | $\mu$ | $\sigma$ | $t_{eh}$ | $t_{eh}$ |
| 5 | | $K_m$ | | | $\gamma$ | $\theta$ | K | $\mu$ | $t_{me}$ | K |
| 6 | | $K_e$ | | | | | | | $t_{em}$ | |
| 7 | | | | | | | | | K | |

proportion of cells in the Venus-positive EpCAM$^{low}$ state were significantly reduced when TGF-β signaling and JAK2/3 were inhibited relative to control groups.

The aforementioned experimental findings suggested that cell-cell communication impacted cell-state-transition rates. However, even though our model set included models with the impact of cell-cell communication on the transition rates, our previous analysis suggested that the $GC_a$&T and $GC_s$I&T models gave the best fit for the spontaneous EMP data. Thus, we investigated whether changes in cell-cell communication via modulation of growth or transition rates improves the fit to the experimental data related to TGF-β and JAK2/3 inhibitor treatment. An alternate scenario can be that treatment with these inhibitors changes the basal cell-state-transition rates. To assess these possibilities, we considered the following three scenarios: (1) optimizing for basal transition rates ($t_{me}$ and $t_{em}$) with inhibitor treatment data while taking best-fit estimates of other parameters (growth/transition influence and carrying capacity) from analysis on spontaneous data, (2) optimizing for growth/transition influence parameter(s) (α and β for $GC_a$&T, σ and μ for $GC_s$I&T, γ for $GC_s$&T-Er, δ for $GC_s$&T-Mr, γ and δ for $GC_s$&T-EMr, and γ and θ for $GC_s$&T-MiEr) with inhibitor treatment data while taking best estimates of other parameters (transition rates and carrying capacity) from analysis on spontaneous data (refer to Tables 4 and 7 for the definition of model specific influence parameter(s) mentioned earlier), and (3) optimizing for both basal transition rates and model-specific growth/transition influence parameter(s) with inhibitor treatment data while obtaining best-fit estimates of other parameters (carrying capacity) from our analysis on spontaneous data.

We averaged the normalized chi-square and AICc values of model fit across inhibitor treatments for a given model and found that, although the lowest averaged normalized chi-square value across population models resulted from scenario 3, the lowest averaged AICc values were obtained for scenario 1 (Figures 3B and S15). We therefore conclude that inhibitor treatment modulated the basal transition rates more significantly than the cell-cell communication (via growth/transition influence parameters), leading to reduced EMT (Figure 3A). Also, the models $GC_a$&T and $GC_s$I&T had either comparable or significantly lower averaged AICc values with respect to other models for the scenario 1 to explain the inhibitor treatment data. Collectively, these findings further assured us that the asymmetric growth competition and GI model with basal transitions ($GC_a$&T and $GC_s$I&T models) are the most probable mechanisms that underlie spontaneous and EMT inhibitor-induced dynamics seen in Figures 2A and 3A.

We next looked at the changes in the basal transition rates $t_{me}$ and $t_{em}$ resulting from the inhibitor treatments (Figures 3C and 3D). The $t_{me}$ rates were significantly larger for post-treatment with anti-TGF-β and AG-490 molecules than for their respective controls (control immunoglobulin G [IgG] and DMSO). On the other hand, $t_{em}$ was reduced, but not significantly for both TGF-β and AG-490 inhibition case (Figure 3C). When comparing the best-fit $t_{me}$ and $t_{em}$ parameters between controls and inhibitor treatment data fitting, we saw 4- and 1.5-fold increases in $t_{me}$ values on treatment with anti-TGF-β and AG-490, respectively, with only slight decrease in the $t_{em}$ values (Figure 3D). Therefore, we conclude that increased M-E transition rates contribute to reduced EMT following inhibitor treatment.

## Analyzing model selection uncertainty via uniqueness of temporal dynamics and sensitivities to experimental data

The best-fit model need not necessarily be closest to the underlying cellular mechanism generating the observed dynamics, given the coupled impact of noise in cellular process with additional noise in experimental measurement. Thus, to quantify how much uncertainty lies in model selection for noisy data as well as what specific experimental conditions a given model is most sensitive to, we performed the following two analyses: (1) cross-model fitting and (2) cross-one validation (leave-one-out)—for models proposed to fit HCC38 cells data.

In cross-model fitting, we checked whether the temporal dynamics resulting from a model, given a parameter set, could be well-explained by any other model in the list (Figure 4A; Table 3). For this, synthetic datasets (having identical structure to the original experimental setup) were generated using randomly sampled parameters for all models in Table 3. For each synthetic dataset generated from a known underlying mechanism (y axis in Figure 4B), every model was used to fit the synthetic data (x axis in Figure 4B), and the procedure was repeated for a total of 1,000 synthetic datasets overall, corresponding to 1,000 independent, randomly sampled parameter sets for each model. After fitting synthetic data from each model to all the models, the AICc were used to identify a model that strikes a trade-off between the goodness of fit and model complexity (number of parameters). Each row of the heatmap (shown in Figure 4B) denotes the percentage distribution of models on the x axis selected by AICc as best model (which has the highest probability to have generated the data) while fitting to the 1,000 synthetic data from a model on y axis. We found that all models performed poorly for being selected as the best model when they fitted other models generated synthetic data with a small amount of noise (refer to columns of Figures 4B and 4I). Exceptions to this trend included the models $GC_s$I&T and $GC_s$&T-EMr which could explain the data generated from other models. However, increasing the level of noise in the synthetic data, in general,

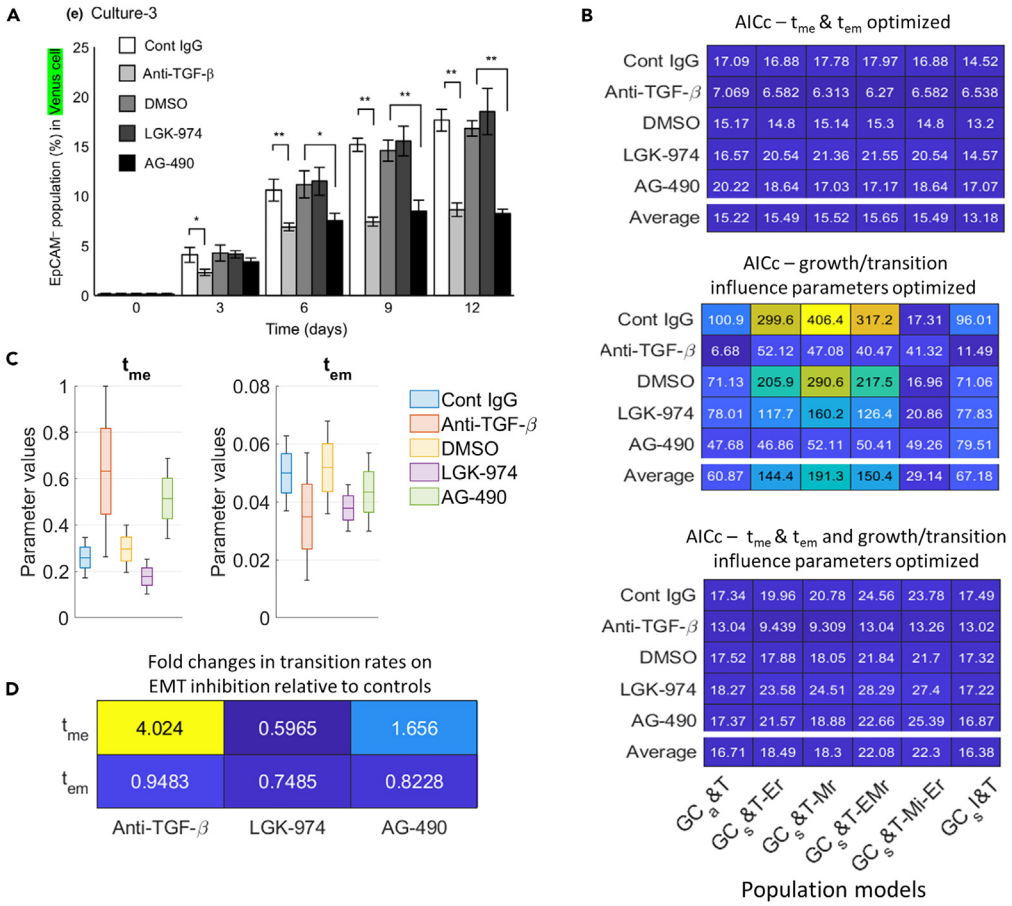



**Figure 3. Changes in basal transition rates better explain the EMT inhibition data than changes in cell-cell communication effects**

(A) Effect of EMT inhibitors on E-M cell-state transitions (adopted from Yamamoto et al., Cancer Science 2017[9]). Here, Cont IgG is the control case for Anti-TGF-β, and DMSO is the control case for both KGK-974 and AG-490.

(B) AICc values of each model upon fitting to the different inhibitor treatment data. Top (resp. middle) panels correspond to scenarios where only basal transition rates (resp. influence parameters) were optimized and the remaining parameters were estimated from the previous analysis of Figure 2 and held constant. Bottom panel corresponds to AICc values when both basal transition rates and influence parameters were optimized. Entries in the bottoms row of each plot represent the average AICc values of their corresponding column.

(C) Distribution of the basal transition rates in $GC_a$&T model across different inhibitor treatment data.

(D) Fold change in the best-fit $t_{me}$ and $t_{em}$ parameters of $GC_a$&T model for the inhibitor treatment data relative to their controls.

led to (1) higher rates of model misidentification (false positives), which was especially true of the $GC_s$&T and $GC_a$&T models as they were selected as the best-fit model for data which they did not generate (columns corresponding to $GC_s$&T vs. $GC_a$&T in Figure 4Bi, ii) and (2) lower rates of correct identification (true positives), wherein the selected best-fit models had actually generated the synthetic data (diagonals of two heatmaps, Figure 4B). Such diminishing effect of increasing noise on model identification was also observed when we analyzed the aforementioned data-generating and data-fitting model pairs for normalized chi-square and *p* value of chi-squre statistic (Figure S16). Thus, increasing noise in the experimental data blurred our framework's ability to identify the underlying mechanism generating the data.

Next, in cross-1-validation analysis, the goal was to check sensitive a model is to initial phenotypic distribution of E and M cells in terms of estimating parameters and recapitulate E-M population dynamics (Figure 5A). For performing this analysis, each culture condition (in panel A) of the experimental data was withheld, one at a time, while fitting the model and estimating the model parameters. The goodness of fit (chi-square values) was then calculated for the holdout culture condition. This fit value was normalized by the control case where goodness of fit (chi-square values) to the same holdout culture condition was calculated when all culture conditions (including the holdout) were used to estimate the parameters. In Figure 5B, the color bar denotes the normalized chi-square values for each holdout condition, and the numerical values denote chi-square values of the respective control case. Here, we see that, although the relative change in goodness of fit is most significant for $GC_s$I&T and $GC_a$&T models with culture condition 3 (initial M-to-E population ratio, M:E = 9:1), only the above models gave a good fit to culture 3 in their respective control cases. Therefore, even if the models $GC_s$I&T and $GC_a$&T were significantly more sensitive to culture condition 3, the absolute change in chi-square values across models can be of similar order. All models were relatively less sensitive to other culture conditions (relative change <20) while

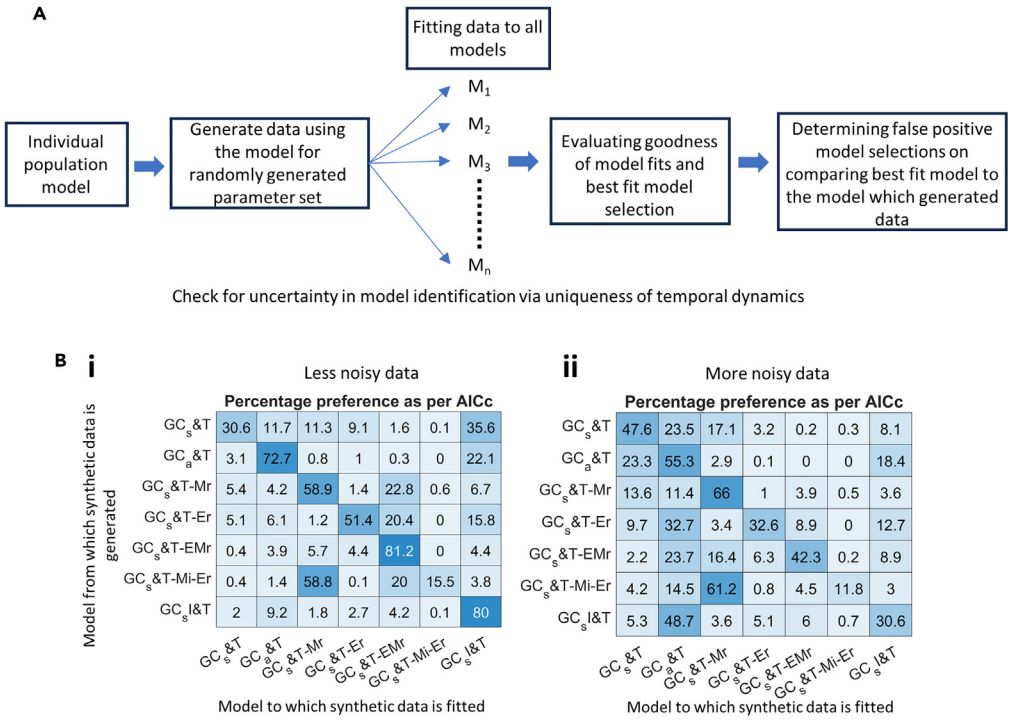



**Figure 4. Analyzing model selection uncertainties via uniqueness of temporal dynamics**

(A) Synthetic data were generated using randomly sampled parameters from each model and then all the models were made to fit the synthetic data. The procedure was repeated for 1,000 synthetic datasets overall for each model.

(B) The values shown in each box of the heatmap denotes the percentage of times a model on the x axis was selected by AICc criteria as the best model while fitting to the 1000 synthetic data from the models on y axis.

also giving good fit to them in their respective controls. Thus, from a modeling standpoint, it is more informative to have experimental data on culture conditions similar to culture 3 (initial 10% E and 90% M proportion) to delineate $GC_a$&T or $GC_s$I&T from other models and further strengthen the conclusion that one of these models capture the true cellular mechanism giving rise to E-M state plasticity.

Since, in both the aforementioned analyses, either we generated synthetic data with structure similar to the experimental data or we used partial experimental data, performing these analyses for very limited PMC42-LA breast cancer data could themselves introduce biases. Therefore, we have not reported the results for the aforementioned two analyses for the PMC42-LA models.

## Adding new experimental data points using the best-fit model to improve confidence bounds on parameters

The best-fit model for HCC38 cell line experimental data ($GC_a$&T) had two practically unidentifiable parameters—β (growth competition to E cells by M cells) and $K_e$ (carrying capacity of E cells)—with an associated poorly defined upper bound and 95% confidence range (Figure 6Ai). For every value of β and $K_e$ below the profile likelihood plot threshold (dashed black line), other model parameters were optimized such that the goodness of fit did not change. This insensitivity of the chi-square values for β and $K_e$ parameter variation propagates to the invariability of temporal trajectories of Venus cells $EpCAM^{low}$ fraction for the optimized parameter sets (Figure S17; columns for β and $K_e$ parameters). However, such invariability of temporal trajectories of Venus $EpCAM^{low}$ fraction to parameter variation is not the case for the other practically identifiable model parameters (Figure S17). Similar variability patterns for the identifiable and unidentifiable model parameters for $GC_s$I&T, G&T, and GI&T models to the (Venus) $EpCAM^{low}$ fractions are shown in Figures S18–S20.

To improve on the parameter bounds of β and $K_e$, we adopted a method discussed by Raue et al. 2009, which identifies a time point of maximum variability in the output (experimental) variable with respect to the parameter of interest, adds new (model-estimated) data corresponding to that time point, and then re-runs the parameter identification analysis on the improved data to check for confidence bound improvement.[34] The overall idea is to improve the characteristics of the existing data to better estimate the model parameters by the addition of new data at the informed time point.

We implemented this algorithm iteratively, i.e., adding a newer data point if previous addition did not improve the bounds on β and $K_e$ parameters until a limit of four additional data points was reached. Given that the output variable (Venus $EpCAM^{low}$ cell fraction) only slightly varied with considerable change in β and $K_e$ parameters, the addition of new data points must be accurate (variation among experimental replicates should be small enough) so that even the slight variation in the output with respect to the aforementioned parameters is captured (Figure S17). Therefore, we saw that the added new data points had very little variation among the replicates (Figure S21B; circles in dark blue

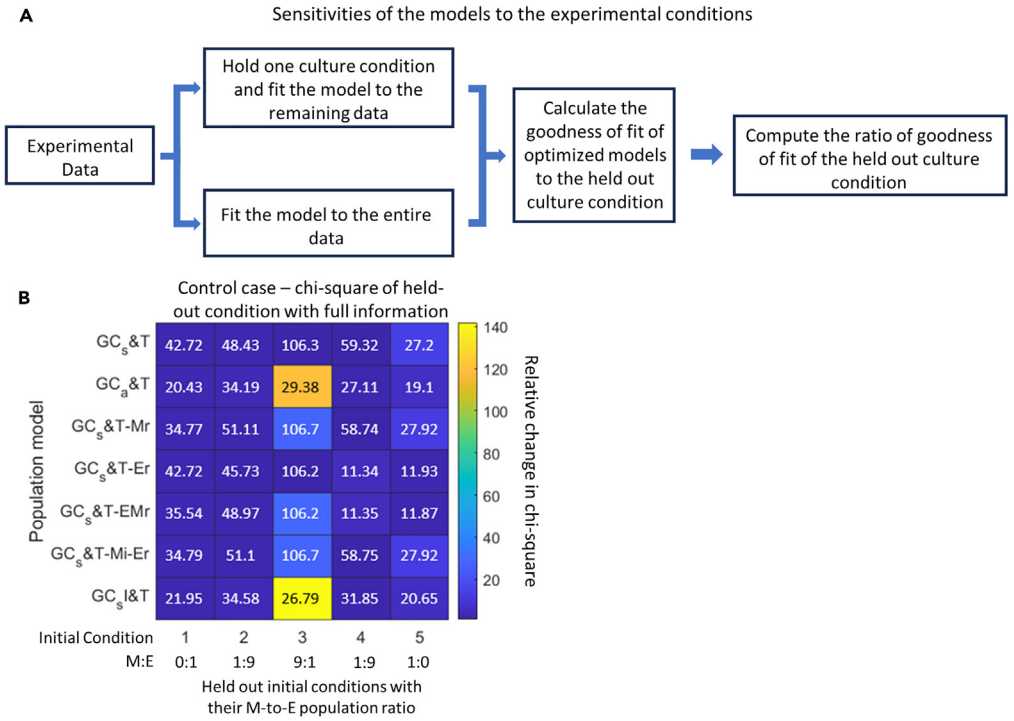



**Figure 5. Analyzing model selection uncertainties via sensitivities to experimental data**

(A) Leave-1-out analysis. Each culture condition was withheld from fitting, one at a time, and then goodness of fit was estimated using the model against the holdout culture condition. The resulting goodness of fit was then normalized by the control case where goodness of fit (chi-square values) to the same heldout culture condition was calculated when all culture conditions (including the one held out) were used to estimate the parameters.

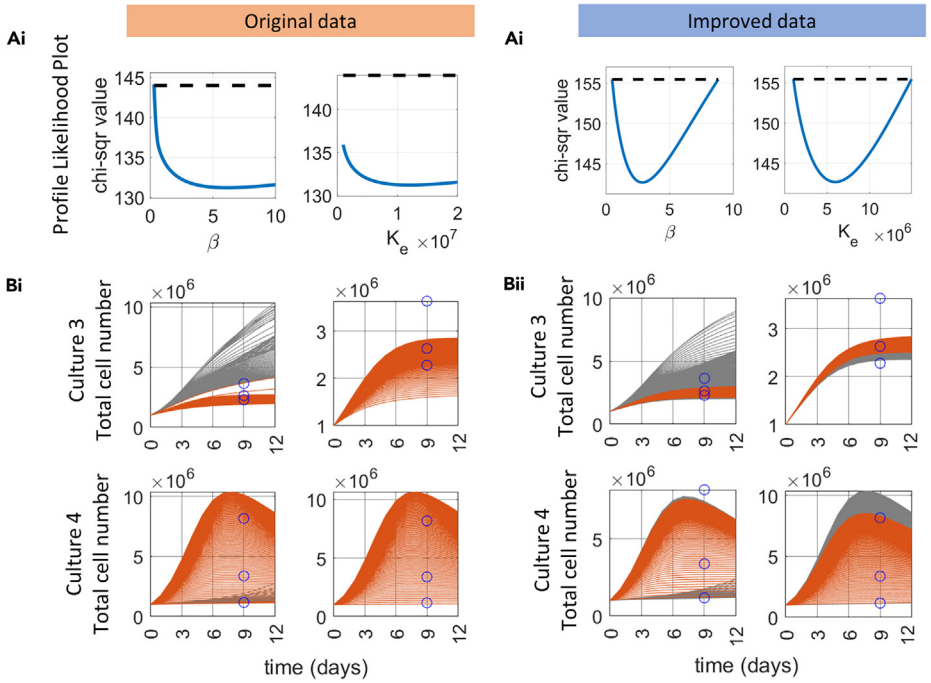
(B) The above calculated relative change in goodness is shown in the color map, and the numerical values in each block denotes the control case goodness of fit (chi-square) values. For leave-one-out analysis, $r_m = (1/54)\ hrs^{-1}$, $r_e = (1/35)\ hrs^{-1}$ and g t ratio = 250; analysis in (A) is agnostic to absolute parameter values if same values/ranges are used for all model's data generation and fitting.

at each new time points). However, the variation among experimental replicates is hard to control, and thus the aforementioned sampling of new data points could not be realized. As total cell number could have been another observable in the experimental data, we next looked into how total cell number varied with the model parameters. All parameters within their 95% confidence bounds showed large variability in the temporal dynamics of total cell numbers, and therefore suggesting quantification of total cell number in the experiment is crucial for precise estimation of model parameters (Figure S22). Therefore, on running the aforementioned algorithm to find the highest variance time point for total cell number for β and $K_e$ parameters, we found that just a single quantification of model-estimated total cell number at day 9 of the experiment, along with experimentally reported Venus EpCAM$^{low}$ fractions, rendered β and $K_e$ parameters to be practically identifiable (Figure 6B). Further, the total cell-number quantification also constrained the 95% confidence ranges of the already identifiable parameters (Figures 6Aii, 6C, S23, and S24). Similar enhanced variability of temporal dynamics of total cell number as compared to Venus EpCAM$^{low}$ cell fraction was observed within the 95% confidence range of $GC_s$I&T model parameters (Figure S25).

## DISCUSSION

Recent experimental and computational analyses have investigated the dynamics of spontaneous and drug-induced cell-state transitions. By fitting a Markov chain model to longitudinal data on the proportion of breast cancer stem and non-stem cell subpopulations, the non-stem-to-stem transition was found to be necessary to explain the experimental data.[36] Cell-state transitions have also been reported among drug-sensitive and drug-resistant cells, wherein either rare drug-resistant cell types exist prior to treatment or cells acquire resistance over the duration of treatment on short-time scales.[2,37–41] It becomes particularly difficult to deconvolute the relative contribution of cell division rates and cell-state transitions in experimental settings that occur over timescales spanning multiple cell divisions. Thus, identifying the mechanisms underlying experimental time course data on the relative frequencies of different interconvertible phenotypes in a cell population remains a formidable task. Nonetheless, mathematical models incorporating cell-state transitions have been able to fit experimental data adequately, thus reinforcing our observations.[39,42,43]

Despite the fact that multiple time course experiments have been performed to capture EMT-associated change in population fractions for many cancer cell lines, a majority of these studies observe state transitions in the presence of an external inducer (growth factor) which confound the cellular regulations that can give rise to spontaneous changeover in the E-M population structure.[6,8,21,44,45] Since experiments




C

| | Parameter | $t_{me}$ ($hr^{-1}$) | $t_{em}$ ($hr^{-1}$) | α | β | $K_m$ (cells) | $K_e$ (cells) |
|---|---|---|---|---|---|---|---|
| Pre-sampling | Lower Bound | 0.00449 | 0.00948 | 0.11005 | 0.31015 | 747073 | 1000000 |
| | Upper bound | 0.0129 | 0.02094 | 4.29714 | 10 | 8452226 | 20000000 |
| Post-sampling | Lower bound | 0.0045 | 0.0119 | 0.1400 | 0.4802 | 1692800 | 1038020 |
| | Upper bound | 0.0114 | 0.0194 | 2.5812 | 8.7423 | 4191500 | 14791400 |

**Figure 6. Adding new, model estimated, total cell numbers experimental data point to improve 95% bounds on parameters**

(A) Profile likelihood plots using original and improved data for parameters β and $K_e$.

(B) Sensitivity analysis for the dynamics of the total cell number under variation in β and $K_e$ values (constrained by threshold (dashed black line) in (Ai) and (Aii), respectively; newly added data points of total cell number are shown in dark blue in (Bii). The red and gray trajectories are drawn using the parameter sets lying within and out of the 95% confidence range, respectively, of the profile likelihood of the focal parameter (A). The sensitivity plots for other parameters and initial cultures are shown in Figures S22 and S24.

(C) Improved 95% confidence range of all model parameters post addition of model estimated total cell number quantification at day 9 to the reported experimental Venus EpCAM$^{low}$ fractions.

performed on PMC42-LA and HCC38 breast cancer cells observed such spontaneous transitions, we specifically considered data from the aforementioned two cell lines in our study.

The cell-nonautonomous (density-dependent) effects seen in the HCC38 cells data initially motivated us to consider population dynamics models with density-dependent cell-state-transition rates ($GC_s$&T-Er, $GC_s$&T-Mr, $GC_s$&T-EMr, and $GC_s$&T-Mi-Er models). However, since the drug-sensitive and drug-resistant subpopulations have also been shown to compete for resources and facilitate and/or suppress the growth of other subpopulations, we then extended our model pool to include the $GC_a$&T and $GC_s$I&T models. Our analysis predicted that density-dependent growth of E and M subpopulations was better than density-dependent transitions in explaining HCC38 breast cancer cells. However, since E-M state transitions play a dominant role in population dynamics on a timescale of few days (Figure 2A), short-term experimental studies fail to capture the long-term contribution of growth interactions to the population dynamics of E-M cells (Figure 2C). Therefore, little is known about how tumor epithelial and mesenchymal cells interact in distinct (resource-rich vs. resource-limited) micro-environments. Answering this question is especially pertinent because of two reasons: (1) the association between E-M phenotypic states and other axes of cell-state plasticity, such as drug-resistance, stemness, and metabolic reprogramming[46–48] and (2) while the resource competition between subpopulations is the fundamental idea underlying adaptive therapy that has already shown promise in delaying the occurrence of resistant in several clinical trials[23,49–51], reciprocal interactions of growth between subpopulations lead to highly aggressive and metastatic disease.[24,27,52,53]

Our results for the $GC_s$I&T model demonstrate that the M cells supress the growth of E cells while E cells facilitate the growth of M cells in density-dependent manner. Thus, we find that, in the resource-limited setting (logistic growth), the asymptotic distribution of the population

tends to be M cell predominant (Figure 2C). What juxtacrine and/or paracrine signaling causes growth-interaction among the PMC42-LA and HCC38 cells is currently unknown; however, the literature presents evidence of other breast cancer cell lines having growth cooperation/suppression among subpopulations through paracrine signaling of metabolites; e.g., lactate and β-hydroxybutyrate secreted by 168FARN murine mammary cells facilitate the growth of another 4T07 murine mammary cells,[28] and estradiol secreted by ribociclib (CDK4/6 inhibitor)-resistant estrogen receptor (ER)-positive cells facilitates the growth of ribociclib-sensitive cells.[27] Further, apart from identification of growth-facilitating factors and quantifying the resultant level of growth upregulation, a few experimental studies have also quantified the resource competition-dependent regulation of growth between the two co-cultured cell populations.[27–29]

Our model also predicts that EMT inhibitors have a larger effect on changes in basal transition rates than they do on cell-cell communication. These predictions are reported in the existing literature across carcinomas. In MDCK cells, mesenchymal phenotype is stabilized via positive feedback in TGF-β/ZEB/miR-200 signaling network. Inhibition of TGF-β signaling with anti-TGF-β antibody led to destabilization of the M state and reduced mesenchymal characteristic of the population.[26] Similarly, in lung cancer cells (A549 and H1650), inhibition of JAK2/STAT3 signaling pathways reduced phosphorylation levels of SMAD3, SNAIL, and MMP2 levels and migration and invasion of TGF-β-induced cells.[54] In NMuMG-EGFR- overexpressed breast cancer cells, EMT facilitated a switch from SRC-dependent EGFR-STAT3 signaling to EGFR-independent Fibronectin-JAK2-STAT3 signaling.[55] Likewise, mesenchymal-like cells in the metastatic cell line MDA-MB-231 had robust activity of fibronectin-JAK2-STAT3 signaling.[55] Thus, both TGF-β and JAK2 signaling are shown to be activated in the mesenchymal state, and their inhibition destabilize mesenchymal state, which in our results is shown by enhanced M-E transition rates ($t_{me}$) (Figure 3C). Nonetheless, from a mathematical model standpoint, the predicted significant role of changes in basal transition rates over cell-cell communication on inhibitor treatment could have also occurred via the following: (1) modulating only cell-cell communication effects in their current simplified forms was insufficient to give the overall fold change required in the effective transition rates ($t_{me}$ and $t_{em}$) to explain the data well of inhibitor treatment and (2) the inhibition of EMT, on blocking of TGF-β with anti-TGF-β antibodies, could have shifted growth dynamics of E-M subpopulations and favored E cells in the experiment, but since the growth rate of E cells was kept constant during our data fitting exercise, the required parametric changes to explain the data well might have reflected into basal transition rates. Additional co-culture experimental data with and without EMT-inhibitor data—in the form of greater numbers of initial culture conditions, temporal data points, or direct experimental estimation of cell growth rates—are necessary to confirm the above competing hypothesis.

The larger ratio of total numbers of experimental data points to the model parameters usually results in well-defined confidence bounds on parameter values. However, if the data do not contain information of certain cellular processes, then even having sufficient time points at which the data are sampled cannot help in bounding parameter values. For example, in the G&T model proposed for fitting E-M transitions in PMC42-LA cells, *a priori* model identifiability analysis predicts that growth rates of E and M cells are not simultaneously identifiable, and this is specific to model cell-fraction dynamics (Data S1; structural identifiability analysis of models proposed for PMC42-LA cells data). Therefore, we need to set either of the growth rates to a constant value to obtain well-defined estimated bounds of the other growth parameter. Thus, the measured variables in the experimental data determine the identifiability of the proposed model parameters. It is for this reason that the only measurement of Venus cell $EpCAM^{low}$ fraction over time in HCC38 cells did not yield well-defined bounds on growth competition parameter (β) and the carrying capacity ($K_e$) of $GC_a$&T model (Figure S21). Consequently, our efforts to improve bounds on β and $K_e$ required measurement of total cell number at day 9 (Figure 6). Extensive computational platforms are being developed for informing future experimental design to ensure that the best set of system variables are captured at the precise time points and model parameter estimates from the data have well-defined confidence bounds.[56–58]

We considered that E and M subpopulation have net growth rates $r_e$ and $r_m$, respectively, which are the difference of division and death rates. The reason for not explicitly accounting for death rates in the proposed models is that, since the changes in the E-M population distribution observed in both PMC42-LA and HCC38 cells data were spontaneous, the effect of enhanced cell death because of growth factor/drug/radiation stimulus is not present and, therefore, contribution from cell death over cell division to the population dynamics is not likely to be significant.[21,53,59] In such a case, the division and death rates both depend on the current population size, and mathematically they can be clubbed to a single net growth rate term. The model could be extended to include explicit death term where its contribution becomes significant.

## Limitations of the study

There are several limitations of this study which are discussed in the following. (1) Limited experimental data present a difficulty to construct a complex model giving insights into the contribution of both density-dependent growth and cell-state transitions to E-M population dynamics. (2) The population-level mathematical models explored here do not exhaustively capture all conceivable mechanisms by which cells can interact with each other. For example, the assumption that E and M fractions modulate growth and/or transition rates in a frequency-dependent manner requires well-mixing among the cell populations and ignores local interactions between neighboring cells that can impact spatial heterogeneity in EMT.[60] The well-mixed case assumption also neglects the concentration gradient of growth factors/cytokines resulting from diffusion in the tissue. (3) The results on E-M population dynamics presented here may deviate from *in vivo* environments where E and M cancer cells compete with other cell types present in the microenvironment that also modulate E-M state-transition rates. (4) Our model is unable to capture the contribution to E-M dynamics coming from the experimentally unreported hybrid E/M and/or stem-non-stem cell substates within the $EpCAM^{low}$ and $EpCAM^{high}$ subpopulations in the PMC42-LA and HCC38 cells data.[13,14,61] Nonetheless, with limited experimental data and different assays, we were able to narrow down to mechanisms that could explain E-M population dynamics observed in PMC42-LA and HCC38 cells and establish the necessity of cell-state transition in determining these dynamics. Overall, our modeling exercise narrows down on the factors that significantly contribute to the E-M population dynamics—(1) spontaneous cell-state transition, (2) density-dependent growth interactions, and (3) changes in basal transition rates on EMT-inhibitor treatment—which could be testified using appropriate design of experiments.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Models
  - Growth (G)
  - Growth influence (GI)
  - Growth and transition (G&T)
  - Growth Influence and Transition (GI&T)
  - Growth and transition – Epithelial retention (G&T-Er)
  - Growth and transition – Mesenchymal retention (G&T-Mr)
  - Growth and transition – Mesenchymal and epithelial retention (G&T-MEr)
  - Growth and transition – Mesenchymal influence and epithelial retention (G&T-Mi-Mr)
  - Growth (symmetric competition)($GC_s$)
  - Growth (symmetric competition) and transition ($GC_s$&T)
  - Growth (asymmetric competition) ($GC_a$)
  - Growth (asymmetric competition) and transition ($GC_a$&T)
  - Growth (symmetric competition) and transition - E retention ($GC_s$&T-Er)
  - Growth (symmetric competition) and transition - M retention ($GC_s$&T-Mr)
  - Growth (symmetric competition) and transition – M & E retention ($GC_s$&T-EMr)
  - Growth (symmetric competition) and transition – M influence E retention ($GC_s$&T-Mi-Er)
  - Growth (symmetric competition), influence ($GC_s$I)
  - Growth (symmetric competition), influence and transition ($GC_s$I&T)
  - Growth (symmetric competition) and transition – Three states ($GC_s$&T3)
  - Growth (symmetric competition) and transition – Three states linear ($GC_s$&T3-L)
  - Parameter identifiability
  - Parameter optimisation
  - Check for the resulting goodness of fit
  - Estimating 95% parameter confidence range and parameter identifiability using profile likelihoods
  - Generating synthetic data
  - Improving identifiability of model parameters by adding model informed new data to the existing experimental data
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110310.

## ACKNOWLEDGMENTS

M.K.J. acknowledges support from the Ramanujan Fellowship (SB/S2/RJN-049/2018) awarded by Science and Engineering Research Board, Department of Science and Technology, Government of India. M.K.J. is also supported by Param Hansa Philanthropies. J.T.G. was supported by the Cancer Prevention and Research Institute of Texas (RR210080). J.T.G. is a CPRIT Scholar in Cancer Research. This research was also supported in part by the International Centre for Theoretical Sciences (ICTS) for participating in the program - Theoretical approaches in cancer progression and treatment (code: ICTS/MATHONCO2024/03).

## AUTHOR CONTRIBUTIONS

P.J., R.K., and M.B.N. performed research. P.J. and R.K. prepared the first manuscript draft. J.T.G. and M.K.J. conceived and supervised the research and obtained funding. All authors contributed to editing the manuscript and analyzed the data.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 7, 2023
Revised: April 21, 2024
Accepted: June 17, 2024
Published: June 19, 2024

## REFERENCES

1. Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor heterogeneity: the Rosetta stone of therapy resistance. Cancer Cell *37*, 471–484. https://doi.org/10.1016/J.CCELL.2020.03.007.
2. Hata, A.N., Niederst, M.J., Archibald, H.L., Gomez-Caraballo, M., Siddiqui, F.M., Mulvey, H.E., Maruvka, Y.E., Ji, F., Bhang, H.E.C., Krishnamurthy Radhakrishna, V., et al. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat. Med. *22*, 262–269. https://doi.org/10.1038/nm.4040.
3. Pillai, M., Hojel, E., Jolly, M.K., and Goyal, Y. (2023). Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. Nat. Comput. Sci. *3*, 301–313. https://doi.org/10.1038/s43588-023-00427-0.
4. Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N., and Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc. Natl. Acad. Sci. USA *110*, 18144–18149. https://doi.org/10.1073/pnas.1318192110.
5. Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A., and Jolly, M.K. (2022). Landscape of epithelial mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. Elife *11*, e76535. https://doi.org/10.7554/ELIFE.76535.
6. Cook, D.P., and Vanderhyden, B.C. (2020). Context specificity of the EMT transcriptional response. Nat. Commun. *11*, 2142–2149. https://doi.org/10.1038/s41467-020-16066-2.
7. Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D'Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013). Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell *154*, 61–74. https://doi.org/10.1016/j.cell.2013.06.005.
8. Panchy, N., Watanabe, K., Takahashi, M., Willems, A., and Hong, T. (2022). Comparative single-cell transcriptomes of dose and time dependent epithelial-mesenchymal spectrums. NAR Genom. Bioinform. *4*, lqac072. https://doi.org/10.1093/NARGAB/LQAC072.
9. Yamamoto, M., Sakane, K., Tominaga, K., Gotoh, N., Niwa, T., Kikuchi, Y., Tada, K., Goshima, N., Semba, K., and Inoue, J.I. (2017). Intratumoral bidirectional transitions between epithelial and mesenchymal cells in triple-negative breast cancer. Cancer Sci. *108*, 1210–1222. https://doi.org/10.1111/cas.13246.
10. Brown, M.S., Abdollahi, B., Wilkins, O.M., Lu, H., Chakraborty, P., Ognjenovic, N.B., Muller, K.E., Jolly, M.K., Christensen, B.C., Hassanpour, S., and Pattabiraman, D.R. (2022). Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. Sci. Adv. *8*. https://doi.org/10.1126/sciadv.abj8002.
11. Bhatia, S., Monkman, J., Blick, T., Duijf, P.H., Nagaraj, S.H., and Thompson, E.W. (2019). Multi-Omics Characterization of the Spontaneous Mesenchymal–Epithelial Transition in the PMC42 Breast Cancer Cell Lines. J. Clin. Med. *8*, 1253. https://doi.org/10.3390/jcm8081253.
12. Bhatia, S., Monkman, J., Blick, T., Pinto, C., Waltham, M., Nagaraj, S.H., and Thompson, E.W. (2019). Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. J. Clin. Med. *8*, 893.
13. Ruscetti, M., Dadashian, E.L., Guo, W., Quach, B., Mulholland, D.J., Park, J.W., Tran, L.M., Kobayashi, N., Bianchi-Frias, D., Xing, Y., et al. (2016). HDAC inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. Oncogene *35*, 3781–3795. https://doi.org/10.1038/onc.2015.444.
14. Biddle, A., Liang, X., Gammon, L., Fazil, B., Harper, L.J., Emich, H., Costea, D.E., and Mackenzie, I.C. (2011). Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative. Cancer Res. *71*, 5317–5326. https://doi.org/10.1158/0008-5472.CAN-11-1059.
15. Zhou, D., Wu, D., Li, Z., Qian, M., and Zhang, M.Q. (2013). Population dynamics of cancer cells with cell state conversions. Quant. Biol. *1*, 201–208. https://doi.org/10.1007/S40484-013-0014-2.
16. Wang, Y., Zhou, J.X., Pedrini, E., Rubin, I., Khalil, M., Taramelli, R., Qian, H., and Huang, S. (2023). Cell population growth kinetics in the presence of stochastic heterogeneity of cell phenotype. J. Theor. Biol. *575*, 111645. https://doi.org/10.1016/J.JTBI.2023.111645.
17. Vega, S., Morales, A.V., Ocaña, O.H., Valdés, F., Fabregat, I., and Nieto, M.A. (2004). Snail blocks the cell cycle and confers resistance to cell death. Genes Dev. *18*, 1131–1143. https://doi.org/10.1101/gad.294104.
18. Comaills, V., Kabeche, L., Morris, R., Buisson, R., Yu, M., Madden, M.W., LiCausi, J.A., Boukhali, M., Tajima, K., Pan, S., et al. (2016). Genomic Instability Is Induced by Persistent Proliferation of Cells Undergoing Epithelial-to-Mesenchymal Transition. Cell Rep. *17*, 2632–2647. https://doi.org/10.1016/j.celrep.2016.11.022.
19. Gollavilli, P.N., Parma, B., Siddiqui, A., Yang, H., Ramesh, V., Napoli, F., Schwab, A., Natesan, R., Mielenz, D., Asangani, I.A., et al. (2021). The role of miR-200b/c in balancing EMT and proliferation revealed by an activity reporter. Oncogene *40*, 2309–2322. https://doi.org/10.1038/S41388-021-01708-6.
20. Granit, R.Z., Masury, H., Condiotti, R., Fixler, Y., Gabai, Y., Glikman, T., Dalin, S., Winter, E., Nevo, Y., Carmon, E., et al. (2018). Regulation of Cellular Heterogeneity and Rates of Symmetric and Asymmetric Divisions in Triple-Negative Breast Cancer. Cell Rep. *24*, 3237–3250. https://doi.org/10.1016/j.celrep.2018.08.053.
21. Devaraj, V., and Bose, B. (2019). Morphological state transition dynamics in EGF-induced epithelial to mesenchymal transition. J. Clin. Med. *8*, 911. https://doi.org/10.3390/jcm8070911.
22. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468. https://doi.org/10.1038/s41586-018-0040-3.
23. Kaznatcheev, A., Peacock, J., Basanta, D., Marusyk, A., and Scott, J.G. (2019). Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. Nat. Ecol. Evol. *3*, 450–456. https://doi.org/10.1038/s41559-018-0768-z.
24. Farrokhian, N., Maltas, J., Dinh, M., Durmaz, A., Ellsworth, P., Hitomi, M., McClure, E., Marusyk, A., Kaznatcheev, A., and Scott, J.G. (2022). Measuring competitive exclusion in non-small cell lung cancer. Sci. Adv. *8*, 7212. https://doi.org/10.1126/SCIADV.ABM7212.
25. Neelakantan, D., Zhou, H., Oliphant, M.U.J., Zhang, X., Simon, L.M., Henke, D.M., Shaw, C.A., Wu, M.F., Hilsenbeck, S.G., White, L.D., et al. (2017). EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. Nat. Commun. *8*, 15773. https://doi.org/10.1038/ncomms15773.
26. Gregory, P.A., Bracken, C.P., Smith, E., Bert, A.G., Wright, J.A., Roslan, S., Morris, M., Wyatt, L., Farshid, G., Lim, Y.-Y., et al. (2011). An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. Mol. Biol. Cell *22*, 1686–1698. https://doi.org/10.1091/mbc.E11-02-0103.
27. Emond, R., Griffiths, J.I., Grolmusz, V.K., Nath, A., Chen, J., Medina, E.F., Sousa, R.S., Synold, T., Adler, F.R., and Bild, A.H. (2023). Cell facilitation promotes growth and survival under drug pressure in breast cancer. Nat. Commun. *14*, 3851. https://doi.org/10.1038/s41467-023-39242-6.
28. Noble, R.J., Walther, V., Roumestand, C., Hochberg, M.E., Hibner, U., and Lassus, P. (2021). Paracrine Behaviors Arbitrate Parasite-Like Interactions Between Tumor Subclones. Front. Ecol. Evol. *9*, 675638. https://doi.org/10.3389/fevo.2021.675638.
29. Freischel, A.R., Damaghi, M., Cunningham, J.J., Ibrahim-Hashim, A., Gillies, R.J., Gatenby, R.A., and Brown, J.S. (2021). Frequency-dependent interactions determine outcome of competition between two breast cancer cell lines. Sci. Rep. *11*, 4908. https://doi.org/10.1038/s41598-021-84406-3.
30. Scheel, C., Eaton, E.N., Li, S.H.J., Chaffer, C.L., Reinhardt, F., Kah, K.J., Bell, G., Guo, W., Rubin, J., Richardson, A.L., and Weinberg, R.A. (2011). Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. Cell *145*, 926–940. https://doi.org/10.1016/j.cell.2011.04.029.
31. Katsuno, Y., Meyer, D.S., Zhang, Z., Shokat, K.M., Akhurst, R.J., Miyazono, K., and

Derynck, R. (2019). Chronic TGF-β exposure drives stabilized EMT, tumor stemness, and cancer drug resistance with vulnerability to bitopic mTOR inhibition. Sci. Signal. *12*, eaau8544. https://doi.org/10.1126/scisignal.aau8544.

32. Hapach, L.A., Wang, W., Schwager, S.C., Pokhriyal, D., Fabiano, E.D., and Reinhart-King, C.A. (2023). Phenotypically sorted highly and weakly migratory triple negative breast cancer cells exhibit migratory and metastatic commensalism. Breast Cancer Res. *25*, 102. https://doi.org/10.1186/s13058-023-01696-3.
33. Dong, R., Goodbrake, C., Harrington, H. A., and Pogudin, G. (2023). Differential elimination for dynamical models via projections with applications to structural identifiability. SIAM Journal on Applied Algebra and Geometry *7*, 194–235. https://doi.org/10.1137/22M1469067.
34. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics *25*, 1923–1929. https://doi.org/10.1093/bioinformatics/btp358.
35. Bellu, G., Saccomani, M.P., Audoly, S., and D'Angiò, L. (2007). DAISY: A new software tool to test global identifiability of biological and physiological systems. Comput. Methods Programs Biomed. *88*, 52–61. https://doi.org/10.1016/J.CMPB.2007.07.002.
36. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644. https://doi.org/10.1016/j.cell.2011.07.026.
37. Li, Q., Wennborg, A., Aurell, E., Dekel, E., Zou, J.Z., Xu, Y., Huang, S., and Ernberg, I. (2016). Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. Proc. Natl. Acad. Sci. USA *113*, 2672–2677. https://doi.org/10.1073/pnas.1519210113.
38. Johnson, K.E., Howard, G.R., Morgan, D., Brenner, E.A., Gardner, A.L., Durrett, R.E., Mo, W., Al'khafaji, A., Sontag, E.D., Jarrett, A.M., et al. (2020). Integrating transcriptomics and bulk time course data into a mathematical framework to describe and predict therapeutic resistance in cancer. Phys. Biol. *18*, 016001. https://doi.org/10.1088/1478-3975/abb09c.
39. Pisco, A.O., Brock, A., Zhou, J., Moor, A., Mojtahedi, M., Jackson, D., and Huang, S. (2013). Non-Darwinian dynamics in therapy-induced cancer drug resistance. Nat. Commun. *4*, 2467. https://doi.org/10.1038/ncomms3467.
40. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435. https://doi.org/10.1038/nature22794.
41. Emert, B.L., Cote, C.J., Torre, E.A., Dardani, I.P., Jiang, C.L., Jain, N., Shaffer, S.M., and Raj, A. (2021). Variability within rare cell states enables multiple paths toward drug resistance. Nat. Biotechnol. *39*, 865–876. https://doi.org/10.1038/s41587-021-00837-3.
42. Zhou, J.X., Pisco, A.O., Qian, H., and Huang, S. (2014). Nonequilibrium population dynamics of phenotype conversion of cancer cells. PLoS One *9*, e110714. https://doi.org/10.1371/journal.pone.0110714.
43. Risom, T., Langer, E.M., Chapman, M.P., Rantala, J., Fields, A.J., Boniface, C., Alvarez, M.J., Kendsersky, N.D., Pelz, C.R., Johnson-Camacho, K., et al. (2018). Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nat. Commun. *9*, 3815. https://doi.org/10.1038/s41467-018-05729-w.
44. Karacosta, L.G., Anchang, B., Ignatiadis, N., Kimmey, S.C., Benson, J.A., Shrager, J.B., Tibshirani, R., Bendall, S.C., and Plevritis, S.K. (2019). Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. Nat. Commun. *10*, 5587. https://doi.org/10.1101/570341.
45. Zhang, J., Tian, X.J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., and Xing, J. (2014). TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci. Signal. *7*, ra91. https://doi.org/10.1126/SCISIGNAL.2005304.
46. Pillai, M., Rajaram, G., Thakur, P., Agarwal, N., Muralidharan, S., Ray, A., Barbhaya, D., Somarelli, J.A., and Jolly, M.K. (2022). Mapping phenotypic heterogeneity in melanoma onto the epithelial-hybrid-mesenchymal axis. Front. Oncol. *12*, 913803. https://doi.org/10.3389/FONC.2022.913803.
47. Sahoo, S., Mishra, A., Kaur, H., Hari, K., Muralidharan, S., Mandal, S., and Jolly, M.K. (2021). A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer *3*, zcab027. https://doi.org/10.1093/NARCAN/ZCAB027.
48. Sahoo, S., Nayak, S.P., Hari, K., Purkait, P., Mandal, S., Kishore, A., Levine, H., and Jolly, M.K. (2021). Immunosuppressive traits of the hybrid epithelial/mesenchymal phenotype. Front. Immunol. *12*, 797261. https://doi.org/10.3389/fimmu.2021.797261.
49. West, J., You, L., Zhang, J., Gatenby, R.A., Brown, J.S., Newton, P.K., and Anderson, A.R.A. (2020). Towards multidrug adaptive therapy. Cancer Res. *80*, 1578–1589. https://doi.org/10.1158/0008-5472.CAN-19-2669.
50. Brady-Nicholls, R., Zhang, J., Zhang, T., Wang, A.Z., Butler, R., Gatenby, R.A., and Enderling, H. (2021). Predicting patient-specific response to adaptive therapy in metastatic castration-resistant prostate cancer using prostate-specific antigen dynamics. Neoplasia *23*, 851–858. https://doi.org/10.1016/j.neo.2021.06.013.
51. Vibishan, B., Harshavardhan, B. V., and Dey, S. (2024). A resource-based mechanistic framework for castration-resistant prostate cancer (CRPC). J. Theor. Biol. *587*, 111806. https://doi.org/10.1016/j.jtbi.2024.111806.
52. Marusyk, A., Tabassum, D.P., Altrock, P.M., Almendro, V., Michor, F., and Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature *514*, 54–58. https://doi.org/10.1038/nature13556.
53. Paczkowski, M., Kretzschmar, W.W., Markelc, B., Liu, S.K., Kunz-Schughart, L.A., Harris, A.L., Partridge, M., Byrne, H.M., and Kannan, P. (2021). Reciprocal interactions between tumour cell populations enhance growth and reduce radiation sensitivity in prostate cancer. Commun. Biol. *4*, 6. https://doi.org/10.1038/s42003-020-01529-5.
54. Liu, R.Y., Zeng, Y., Lei, Z., Wang, L., Yang, H., Liu, Z., Zhao, J., and Zhang, H.T. (2014). JAK/STAT3 signaling is required for TGF-β-induced epithelial-mesenchymal transition in lung cancer cells. Intern. J. Oncol. *44*, 1643–1651. https://doi.org/10.3892/ijo.2014.2310.
55. Balanis, N., Wendt, M.K., Schiemann, B.J., Wang, Z., Schiemann, W.P., and Carlin, C.R. (2013). Epithelial to mesenchymal transition promotes breast cancer progression via a fibronectin-dependent STAT3 signaling pathway. J. Biol. Chem. *288*, 17954–17967. https://doi.org/10.1074/jbc.M113.475277.
56. Chhajer, H., and Roy, R. (2023). Rationalised experiment design for parameter estimation with sensitivity clustering. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.11.561860.
57. Franceschini, G., and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. Chem. Eng. Sci. *63*, 4846–4872. https://doi.org/10.1016/j.ces.2007.11.034.
58. Beik, S.P., Harris, L.A., Kochen, M.A., Sage, J., Quaranta, V., and Lopez, C.F. (2023). Unified tumor growth mechanisms from multimodel inference and dataset integration. PLoS Comput. Biol. *19*, e1011215. https://doi.org/10.1371/JOURNAL.PCBI.1011215.
59. Howard, G.R., Jost, T.A., Yankeelov, T.E., and Brock, A. (2022). Quantification of long-term doxorubicin response dynamics in breast cancer cell lines to direct treatment schedules. PLoS Comput. Biol. *18*, e1009104. https://doi.org/10.1371/journal.pcbi.1009104.
60. Boareto, M., Jolly, M.K., Goldman, A., Pietilä, M., Mani, S.A., Sengupta, S., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2016). Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. J. R. Soc. Interface *13*, 20151106. https://doi.org/10.1098/rsif.2015.1106.
61. Biddle, A., Gammon, L., Liang, X., Costea, D.E., and Mackenzie, I.C. (2016). Phenotypic Plasticity Determines Cancer Stem Cell Therapeutic Resistance in Oral Squamous Cell Carcinoma. EBioMedicine *4*, 138–145. https://doi.org/10.1016/J.EBIOM.2016.01.007.
62. Taylor, L.R., Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1987). Numerical Recipes: The Art of Scientific Computing. J. Anim. Ecol. *56*, 374. https://doi.org/10.2307/4830.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Analysed PMC42-LA Cells Data | Bhatia et al.[12] | https://github.com/Paras-Jain20/EMT-Cell-non-autonomus-interactions/tree/main/Experimetal%20Data |
| Analysed HCC38 Cells Data | Yamamoto et al.[9] | https://github.com/Paras-Jain20/EMT-Cell-non-autonomus-interactions/tree/main/Experimetal%20Data |
| Software and algorithms | | |
| DAISY (for Structural Identifiability) | Bellu et al.[35] | DAISY (unipd.it) |
| Differential algebraic elimination in Julia (for Structural Identifiability) | Dong et al.[33] | SciML/StructuralIdentifiability.jl: Fast and automatic structural identifiability software for ODE systems (github.com) |
| Non-linear least squares fitting | MathWorks | Solve nonlinear least-squares (nonlinear data-fitting) problems - MATLAB lsqnonlin - MathWorks India) |
| Profile Likelihood Analysis | This paper | |
| Improving Practical Identifiability algorithm | This paper | |
| Cross-model fitting algorithm | This paper | |
| Hold-one out analysis algorithm | This paper | |

## RESOURCE AVAILABILITY

### Lead contact

For additional resources or inquiries please kindly contact Prof. Mohit Kumar Jolly (mkjolly@iisc.ac.in).

### Materials availability

This paper did not include newly generated reagents.

### Data and code availability

- The data used in this paper are publicly available.
- All data and scripts used in the manuscript are uploaded and available at https://github.com/Paras-Jain20/EMT-Cell-non-autonomus-interactions.
- For further details needed to re-examine the data analysis and methodology presented in this paper, please kindly request the authors for additional information.

## METHOD DETAILS

### Models

Table 1 represents the models that are proposed to fit spontaneous EMP seen in PMC42-LA breast cancer cells.[12] All models consider exponential cell growth, with growth rate of the M subpopulation ($r_m$) set to a constant value and all rate parameters (E-growth, E-M and M-E transition rates) being scaled by M cells growth rates at the time of parameter optimization. Further, we constrained the growth rate of the E subpopulation ($r_e$) to be greater than that of M subpopulation ($r_m$) as reported experimentally,[17,18] giving a normalised $r_e \geq 1$. The temporal dynamics is realised in terms changes in E and M cell fractions in the population, rather than changes in E and M cell numbers. Following are the mathematical description of each model:

### Growth (G)

Consider *E* and *M* denotes epithelial and mesenchymal cell numbers which are growing at per-capita rate of $r_e$ and $r_m$ respectively,

$$\frac{dM}{dt} = r_m M$$

$$\frac{dE}{dt} = r_e E$$

Calculating the dynamics for the epithelial and mesenchymal fraction, $e = E/(E+M)$ and $m = M/(E+M)$:

$$\frac{d\left(\frac{E}{E+M}\right)}{dt} = \frac{(E+M)\frac{dE}{dt} - E\frac{d(E+M)}{dt}}{(E+M)^2}$$

$$\frac{d\left(\frac{E}{E+M}\right)}{dt} = r_e\, e - \frac{e}{E+M}\,(r_m E + r_e M)$$

$$\frac{de}{dt} = r_e e - (r_m m + r_e e)\, e$$

Similarly, dynamic equation for mesenchymal fraction ($m$) can be obtained:

$$\frac{dm}{dt} = r_m m - (r_m m + r_e e)\, m$$

Where, $m$ and $e$ are mesenchymal and epithelial cell fraction in the population, and $r_m$ and $r_e$ are the M and E cells growth rates, respectively. The product of $(r_m m + r_e e)$ term with $m$ in first equation and with $e$ in the second equation above comes because of the conversion of cell number dynamics to cell-fraction dynamics.

The following description of models shows the dynamics of E and M cell-fractions (which is obtained following the above shown conversion).

### Growth influence (GI)

$$\frac{dm}{dt} = r_m(1 - \sigma e)m - (r_m(1 - \sigma e)m + r_e(1 - \mu m)e)\, m$$

$$\frac{de}{dt} = r_e(1 - \mu m)e - (r_m(1 - \sigma e)m + r_e(1 - \mu m)e)\, e$$

Where, $\sigma$ and $\mu$ are the influence of E cells on M cells' growth and M cells influence on E cell's growth. The range of $\sigma$ and $\mu$ during parameter optimization is set between [-1, 1].

### Growth and transition (G&T)

$$\frac{dm}{dt} = (r_m - t_{me})m + t_{em}e - (r_m m + r_e e)\, m$$

$$\frac{de}{dt} = (r_e - t_{em})e + t_{me}m - (r_m m + r_e e)\, e$$

Where, $t_{me}$ and $t_{em}$ are the per-capita M-E and E-M cell-transition rates.

### Growth Influence and Transition (GI&T)

$$\frac{dm}{dt} = r_m\,(1 - \alpha e)m - t_{me}m + t_{em}e - (r_m(1 - \sigma e)m + r_e(1 - \mu m)e)\, m$$

$$\frac{de}{dt} = r_e(1 - \beta m)e - t_{em}e + t_{me}m - (r_m(1 - \sigma e)m + r_e(1 - \mu m)e)\, e$$

### Growth and transition – Epithelial retention (G&T-Er)

$$\frac{dm}{dt} = (r_m - t_{me})m + t_{em}(1 - \gamma e)e - (r_m m + r_e e)\, m$$

$$\frac{de}{dt} = (r_e - t_{em}(1 - \gamma e))e + t_{me}m - (r_m m + r_e e)\, e$$

Where, $\gamma$ sets the strength of E cells to reduce their E-M transition rates. The range of $\gamma$ is set between [0,1] during optimization.

### Growth and transition – Mesenchymal retention (G&T-Mr)

$$\frac{dm}{dt} = (r_m - t_{me}(1 - \delta m))m + t_{em}e - (r_m m + r_e e)\, m$$

$$\frac{de}{dt} = (r_e - t_{em})e + t_{me}(1 - \delta m)m - (r_m m + r_e e)\, e$$

Where, parameter $\delta$ sets strength of M cells to reduce their M-E transition rates. The range of $\theta$ is set between [0,1] during optimization.

### Growth and transition – Mesenchymal and epithelial retention (G&T-MEr)

$$\frac{dm}{dt} = (r_m - t_{me}(1 - \delta m))m + t_{em}(1 - \gamma e)e - (r_m m + r_e e)\, m$$

$$\frac{de}{dt} = (r_e - t_{em}(1 - \gamma e))e + t_{me}(1 - \delta m)m - (r_m m + r_e e)\, e$$

### Growth and transition – Mesenchymal influence and epithelial retention (G&T-Mi-Mr)

$$\frac{dm}{dt} = (r_m - t_{me}(1 - \delta m))m + t_{em}(1 + \theta m)e - (r_m m + r_e e)\, m$$

$$\frac{de}{dt} = (r_e - t_{em}(1 + \theta m))e + t_{me}(1 - \delta m)m - (r_m m + r_e e)\, m$$

Where, $\theta$ sets the strength of M cells to enhance the E-M transition rates. The range of $\theta$ is set between [1,4] during optimization.

Table 4 represents the models proposed to fit spontaneous EMP seen in HCC38 breast cancer cells.[9] All models consider logistic growth of cells, with all rate parameters (E-growth, E-M and M-E transition rates) being scaled by M cells growth rates at the time of parameter optimization. Further, the temporal dynamics is realised in terms of changes cell numbers of four subpopulations – 1) Venus non-labelled EpCAM$^{low}$ cells, 2) Venus non-labelled EpCAM$^{high}$ cells, 3) Venus labelled EpCAM$^{low}$ cells, and 4) Venus labelled EpCAM$^{high}$ cells. The characteristics (growth, resource competition, transition, and cell-cell influence) of Venus labelled and non-labelled population for a given EpCAM status are kept same. Following are the mathematical description of each model:

### Growth (symmetric competition)($GC_s$)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right)$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right)$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right)$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right)$$

Where, $E_{VP/VN}$ and $M_{VP/VN}$ are EpCAM$^{high}$ and EpCAM$^{low}$ cell numbers with subscript VN and VP denoting their Venus positive and Venus negative status. $r_e$ and $r_m$ are growth rates of E (EpCAM$^{high}$) cells and M (EpCAM$^{low}$) cells, respectively. K is the carrying capacity (maximum population size). Total E cells, $E = E_{VN} + E_{VP}$, and total M cells, $M = M_{VN} + M_{VP}$.

### Growth (symmetric competition) and transition ($GC_s$&T)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VN} + t_{em}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VN} - t_{em}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VP} + t_{em}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VP} - t_{em}E_{VP}$$

Where, $t_{me}$ and $t_{em}$ are the per-capita M-E and E-M cell-transition rates.

### Growth (asymmetric competition) ($GC_a$)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(\alpha(E_{VN}+E_{VP})+M_{VN}+M_{VP})}{K_m}\right)$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+\beta(M_{VN}+M_{VP}))}{K_e}\right)$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(\alpha(E_{VN}+E_{VP})+M_{VN}+M_{VP})}{K_m}\right)$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+\beta(M_{VN}+M_{VP}))}{K_e}\right)$$

Where, $\alpha$ and $\beta$ parameters capturing E cell's influence on M cells growth and M cell's influence on E cell's growth. The range of both $\alpha$ and $\beta$ parameters is in [0 10] during the parameter optimization. $K_e$ and $K_m$ are the respective carrying capacity of E and M cells.

### Growth (asymmetric competition) and transition ($GC_a$&T)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(\alpha(E_{VN}+E_{VP})+M_{VN}+M_{VP})}{K_m}\right) - t_{me}M_{VN} + t_{em}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+\beta(M_{VN}+M_{VP}))}{K_e}\right) + t_{me}M_{VN} - t_{em}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(\alpha(E_{VN}+E_{VP})+M_{VN}+M_{VP})}{K_m}\right) - t_{me}M_{VP} + t_{em}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+\beta(M_{VN}+M_{VP}))}{K_e}\right) + t_{me}M_{VP} - t_{em}E_{VP}$$

### Growth (symmetric competition) and transition - E retention ($GC_s$&T-Er)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VN} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VN} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VP} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VP} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VP}$$

Where, $\gamma$ sets the strength of E cells ($E = E_{VN} + E_{VP}$) to reduce their E-M transition rates. The range of $\gamma$ is set between [0,1] during optimization.

### Growth (symmetric competition) and transition - M retention ($GC_s$&T-Mr)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VN} + t_{em}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VN} - t_{em}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VP} + t_{em}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VP} - t_{em}E_{VP}$$

Where, $\delta$ sets the strength of M cells ($M = M_{VN} + M_{VP}$) to reduce their M-E transition rates. The range of $\delta$ is set between [0,1] during optimization.

### Growth (symmetric competition) and transition – M & E retention ($GC_s$&T-EMr)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VN} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VN} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VP} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}\left(1 - \delta\left(\frac{M}{E+M}\right)\right)M_{VP} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right)\right)E_{VP}$$

### Growth (symmetric competition) and transition – M influence E retention ($GC_s$&T-Mi-Er)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VN} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right) + \theta\left(\frac{M}{E+M}\right)\right)E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VN} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right) + \theta\left(\frac{M}{E+M}\right)\right)E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) - t_{me}M_{VP} + t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right) + \theta\left(\frac{M}{E+M}\right)\right)E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K}\right) + t_{me}M_{VP} - t_{em}\left(1 - \gamma\left(\frac{E}{E+M}\right) + \theta\left(\frac{M}{E+M}\right)\right)E_{VP}$$

Where, $\theta$ sets the strength of M cells to enhance the E-M transition rates. The range of $\theta$ is set between [1,4] during optimization.

### Growth (symmetric competition), influence ($GC_s$I)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \sigma\frac{(E_{VN}+E_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right)$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \mu\frac{(M_{VN}+M_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right)$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \sigma\frac{(E_{VN}+E_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right)$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \mu\frac{(M_{VN}+M_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right)$$

### Growth (symmetric competition), influence and transition ($GC_s$I&T)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \sigma\frac{(E_{VN}+E_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right) - t_{me}M_{VN} + t_{em}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \mu\frac{(M_{VN}+M_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right) + t_{me}M_{VN} - t_{em}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \sigma\frac{(E_{VN}+E_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right) - t_{me}M_{VP} + t_{em}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}{K} - \mu\frac{(M_{VN}+M_{VP})}{(E_{VN}+E_{VP}+M_{VN}+M_{VP})}\right) + t_{me}M_{VP} - t_{em}E_{VP}$$

### Growth (symmetric competition) and transition – Three states ($GC_s$&T3)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{N}{K}\right) - t_{me}M_{VN} - t_{mh}M_{VN} + t_{em}E_{VN} + t_{hm}H_{VN}$$

$$\frac{dH_{VN}}{dt} = r_h H_{VN}\left(1 - \frac{N}{K}\right) - t_{hm}H_{VN} - t_{he}H_{VN} + t_{mh}M_{VN} + t_{eh}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{N}{K}\right) + t_{me}M_{VN} + t_{he}H_{VN} - t_{em}E_{VN} - t_{eh}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{N}{K}\right) - t_{me}M_{VP} - t_{mh}M_{VP} + t_{em}E_{VP} + t_{hm}H_{VP}$$

$$\frac{dH_{VP}}{dt} = r_h H_{VP}\left(1 - \frac{N}{K}\right) - t_{hm}H_{VP} - t_{he}H_{VP} + t_{mh}M_{VP} + t_{eh}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{N}{K}\right) + t_{me}M_{VP} + t_{he}H_{VP} - t_{em}E_{VP} - t_{eh}E_{VP}$$

$$N(t) = E_{VN}(t) + E_{VP}(t) + H_{VN}(t) + H_{VP}(t) + + M_{VN}(t) + M_{VP}(t)$$

Where, H and h denote the hybrid states; $t_{me}$, $t_{mh}$, $t_{hm}$, $t_{he}$, $t_{em}$, and $t_{eh}$ are the per-capita cell-transition rates.

### Growth (symmetric competition) and transition – Three states linear (GC$_s$&T3-L)

$$\frac{dM_{VN}}{dt} = r_m M_{VN}\left(1 - \frac{N}{K}\right) - t_{mh}M_{VN} + t_{hm}H_{VN}$$

$$\frac{dH_{VN}}{dt} = r_h H_{VN}\left(1 - \frac{N}{K}\right) - t_{hm}H_{VN} - t_{he}H_{VN} + t_{mh}M_{VN} + t_{eh}E_{VN}$$

$$\frac{dE_{VN}}{dt} = r_e E_{VN}\left(1 - \frac{N}{K}\right) + t_{he}H_{VN} - t_{eh}E_{VN}$$

$$\frac{dM_{VP}}{dt} = r_m M_{VP}\left(1 - \frac{N}{K}\right) - t_{mh}M_{VP} + t_{hm}H_{VP}$$

$$\frac{dH_{VP}}{dt} = r_h H_{VP}\left(1 - \frac{N}{K}\right) - t_{hm}H_{VP} - t_{he}H_{VP} + t_{mh}M_{VP} + t_{eh}E_{VP}$$

$$\frac{dE_{VP}}{dt} = r_e E_{VP}\left(1 - \frac{N}{K}\right) + t_{he}H_{VP} - t_{eh}E_{VP}$$

$$N(t) = E_{VN}(t) + E_{VP}(t) + H_{VN}(t) + H_{VP}(t) + + M_{VN}(t) + M_{VP}(t)$$

Where, H and h denote the hybrid states; $t_{mh}$, $t_{hm}$, $t_{he}$, and $t_{eh}$ are the per-capita cell-transition rates.

### Parameter identifiability

Parameters can be classified as:

(1) Locally Identifiable: the value can be recovered up to finitely many options;
(2) Globally Identifiable: the value can be recovered uniquely.

We checked for *a priori* global identifiability of parameters using a using Differential algebra approach. For the model proposed to fit PMC42-LA experimental data (Table 1) we performed global identifiability calculations analytically (Data S1- Structural Identifiability analysis of models proposed for PMC42-LA cells data). However, for the more complex models proposed for HCC38 experimental data (Table 4) we used DAISY and StructuralIdentifiability.jl package.[33,35]

A posteriori local identifiability of model parameter estimates was performed using Profile Likelihood analysis.[34]

### Parameter optimisation

Parameter optimization involves the following steps.

1. Generating trajectories by simulating the population dynamics using the proposed models.
2. Quantifying the difference between the simulated data and the experimental data using a cost/objective (goodness of fit) function. We used the chi-square ($\chi^2$) values as the cost function:

$$\chi^2 = \sum_{culture=1}^{M}\sum_{rep=1}^{N}\sum_{obs=1}^{P}\left(\frac{x_{obs}^{data} - x_{obs}^{model}}{\sigma_{obs}^{data}}\right)^2$$

Where, *obs* refer to each experimental observation/measurements, $x_{obs}^{data}$, for a given replicate and culture condition with total number of P measurements within the duration of the experiment. The difference between temporal measurements of the experimental data and model simulation is summed for each measurement time point, replicate (rep) and culture condition (culture). $\sigma_{obs}^{data}$ is the standard deviation among the replicates for each measurement time point and the culture condition. M, N and P are number of culture conditions, replicates, and measurement time points, respectively. M = 2, N = 3, and P = 4 for PMC42-LA cells and M = 5, N = 3, and P = 5 for HCC38 cells spontaneous EMP data; and M = 1, N = 3, and P = 5 for HCC38 cells inhibitor treatment data.

3. Searching for the parameter set that minimizes the objective function using an optimizer. The parameter optimization was carried out using the 'lsqnonlin' function from MATLAB's Optimisation toolbox. By default, it uses the 'trust region reflective' algorithm, in order to converge to the minimum value of the cost function. The default values of optional parameter of 'lsqnonlin' were used for parameter estimation.

## Check for the resulting goodness of fit

An optimized model with a minimal chi-square value relative to the other models does not convey whether the optimized model is actually giving a good fit to the data, or even overfitting the data. Such uncertainty in goodness of fit by least chi-square values comes because there is no reference point to which these values should be compared to. Therefore, to ensure a good model fit occurs concomitantly with the model giving minimal chi-squares, we further analysed:

(1) Normalized chi-square values - We normalize the chi-square values of the model fits to the number of data points. A normalized chi-square value close to 1 indicates that all the experiment data points are around one standard deviation away from fitted data point.[62]
(2) p-value of chi-square statistic – Assuming variations in the data are normally distributed and the model is linear in its parameters, the resulting chi-square values follow chi-square probability distribution with N-M degrees of freedom. Here, N is the total number of data points and M is the number of fitted parameters. Therefore, we can determine the probability that a chi-square value will exceed the model obtained chi-square value by random chance. The closer the above probability is to a value of 1, the greater certainty there is that the model is reliably capturing meaningful trends in the experimental data.[62]

## Estimating 95% parameter confidence range and parameter identifiability using profile likelihoods

We performed the Profile Likelihood (PL) analysis to estimate the 95% confidence range of optimized model parameters.[34] The PL analysis uses chi-square statistics to define a threshold level of change in chi-square values from the minimum chi-square values under which model fits are equally good (cost function/goodness of fit). This permissible change in the chi-square from the point of minima depends upon number of model parameters (degree of freedom), and is used to define 95% confidence range for each model parameter. To generate profile likelihood plot of a particular parameter (say $\theta_i$) in a model of N total parameters ($\theta_{1,\ldots,N}$),

1. We first choose a relevant range of the parameter ($\theta_i$).
2. Discretize the above parametric range into 1000 values separated by equal steps. Let's define each value in the discretized interval by $\theta_i^k$, where k = 1 to 1000.
3. Incrementally for each discretized value of $\theta_i$ in its range ($\theta_i^k$), we optimize the other N-1 parameters ($\theta^k_{1,\ldots,i-1,i+1,\ldots,N}$) of the model to minimize the chi-square value while keeping the value of $\theta_i^k$ constant.

$$\chi^2_{PL}\left(\theta_i^k\right) = \min_{\theta_{j\neq i}}\left[\chi^2\left(\theta_j^k\right)\right]$$

4. The minimum chi-square value thus obtained at each of these points corresponds to the profile likelihood value at the point.
5. While carrying out the optimization for increasing $\theta_i^k$ values ($k^{th}$ step), the input guess values for other N-1 parameters ($\theta^k_{1,\ldots,i-1,i+1,\ldots,N}$) given to the optimization algorithm are taken from the optimized parameter sets for the last step $\theta_i^{k-1}$ value ((k-1)$^{th}$ step), i.e., $\theta^k_{1,\ldots,i-1,i+1,\ldots,N} = \underset{\theta_{j\neq i}}{\operatorname{argmin}}\,[\chi^2(\theta_j^{k-1})]$.

## Generating synthetic data

To generate synthetic data for a given modal:

(1) Multiple parameters set of the model were generated by uniformly sampling each parameter within its lower and upper bounds.
(2) For each parameter set, the model dynamics were simulated using the experimental culture condition as the starting point.
(3) We then sampled the temporal dynamics of the model output (EpCAM$^{low}$ fraction in models proposed for PMC42-LA cells data, and Venus Cells EpCAM$^{low}$ fraction in models proposed for HCC38 cells data) at limited time points, mimicking the experimental measurement design.

(4) Since, for each parameter set for given a model and culture condition we can generate only one temporal trajectory, we added normally distributed noise to the sampled data at each timepoints to get three independent samples, thus mimicking experimental replicates for the culture condition. The standard deviation of normalised distributed noise was adjusted using a denominator term 'noise factor'. Thus, increasing 'noise factor' reduced the variation among replicates of the synthetic data. Further, in cases where adding noise led to negative output (negative E and M fraction) values then resampling of noise term was performed until a positive sum of data with noise term was obtained.
(5) We gave same structure to the synthetic data as with the original experimental data for models for HCC38 cells data, i.e., same culture conditions, output measurement timepoints, and number of replicates. For synthetic data generated with models proposed for PMC42-LA cells, the number of measurements timepoints where increased from four to eight, and the culture conditions and replicates were same as the experimental data.

### Improving identifiability of model parameters by adding model informed new data to the existing experimental data

Following are the steps to iteratively identify a timepoint of maximum variability in the output (experimental) variable with respect to parameter of interest and add new (model estimated) data to that time point:

1. For a non-identifiable parameter of a model, we simulate the model dynamics for all parameter sets that are present within the 95% confidence intervals for the non-identifiable parameter, starting with E and M proportions as in the original culture conditions. Note that each point/step in the PL analysis of a parameter corresponds to a set of values of all model parameters. This results in output (Venus cells $EpCAM^{low}$ fraction) variation due to parametric variation. The standard deviation of the output for each culture conditions is calculated at uniformly spaced time points and the standard deviations across culture conditions are then summed up time point wise to give a single time series of standard deviation values for the non-identifiable parameter.
2. The above step is repeated for all non-identifiable parameters of the model to obtain 'n' distinct time series of summed standard deviations for the 'n' non-identifiable parameters.
3. From the set of 'n' distinct time series obtained in step 2, we then find a time point, and the parameter index across 'n' time series which has the largest standard deviation values. If the selected time point is already present in the experimental data, then we find another combination of time point and parameter index which results in second (next) largest standard deviation value across 'n' time series. The selected time point is the optimal time at which new data will be added using the output variation observed in the selected parameter index.
4. In the output variation observed for each culture condition of the selected parameter index in the step 3, we sample three new output values at the optimal time point. These sampled new values corresponding to the optimal time point for each culture condition are then added to the existing experimental data.
5. With the updated experimental data run the PL analysis for all non-identifiable parameters of step 1 and 2. If any one of the parameters remains non-identifiable then go back to step 1, otherwise stop. The addition of new data is stopped even when a maximum of 4 new data are already added to the original experimental data.

## QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical tests were performed in the study.