

# Population heterogeneity in the epithelial to mesenchymal transition is controlled by NFAT and phosphorylated Sp1

Russell Gould<sup>1</sup>, Anirikh Chakrabarti<sup>2</sup>, Jeffrey D. Varner<sup>2</sup>, Jonathan Butcher<sup>1,\*</sup>,  
**1 Department of Biomedical Engineering, Cornell University, Ithaca, NY, USA**

**2 School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA**

\* E-mail: jtb47@cornell.edu

## Abstract

Epithelial to mesenchymal transition (EMT) is an essential differentiation program during tissue morphogenesis and remodeling. EMT is induced by soluble transforming growth factor  $\beta$  (TGF- $\beta$ ) family members, and restricted by vascular endothelial growth factor family members. While many downstream molecular regulators of EMT have been identified, these have been largely evaluated individually without considering potential crosstalk. In this study, we created an ensemble of dynamic mathematical models describing TGF- $\beta$  induced EMT to better understand the operational hierarchy of this complex molecular program. These models incorporate mass action kinetics within an ordinary differential equation (ODE) framework to describe the transcriptional and post-translational regulatory events driving EMT. Model parameters were estimated from multiple data sets using multiobjective optimization, in combination with cross-validation. TGF- $\beta$  exposure drove the model population toward a mesenchymal phenotype, while an epithelial phenotype was maintained following vascular endothelial growth factor A (VEGF-A) exposure. Simulations predicted that the transcription factors phosphorylated SP1 and NFAT were master regulators promoting or inhibiting EMT, respectively. Surprisingly, simulations also predicted that a cellular population could exhibit phenotypic heterogeneity (characterized by a significant fraction of the population with both high epithelial and mesenchymal marker expression) if treated simultaneously with TGF- $\beta$  and VEGF-A. We tested this prediction experimentally in both MCF10A and DLD1 cells and found that upwards of 45% of the cellular population acquired this hybrid state in the presence of both TGF- $\beta$  and VEGF-A. We experimentally validated the predicted NFAT/Sp1 signaling axis for each phenotype response. Lastly, we found that cells in the hybrid state had significantly different functional behavior when compared to VEGF-A or TGF- $\beta$  treatment alone. Together, these results establish a predictive mechanistic model of EMT susceptibility, and potentially reveal a novel signaling axis which regulates carcinoma progression through an EMT versus tubulogenesis response.

## Author Summary

Tissue formation and remodeling requires a complex and dynamic balance of interactions between epithelial cells, which reside on the surface, and mesenchymal cells that reside in the tissue interior. During embryonic development, wound healing, and cancer, epithelial cells transform into a mesenchymal cell to form new types of tissues. It is important to understand this process so that it can be controlled to generate beneficial effects and limit pathological differentiation. Much research over the past 20 years has identified many different molecular species that are relevant, but these have mainly been studied one at a time. In this study, we developed and implemented a novel computational strategy to interrogate all of the known players in this transformation process to identify which are the major bottlenecks. We determined that NFATc1 and pSP1 are essential for promoting epithelial or mesenchymal differentiation, respectively. We then predicted the existence of a partially transformed cell that exhibits both epithelial and mesenchymal characteristics. We found this partial cell type develops a network of invasive but stunted vascular structures that may be a unique cell target for understanding cancer progression and angiogenesis.

## Introduction

The epithelial to mesenchymal transition (EMT) is a broadly participating, evolutionarily conserved differentiation program essential for tissue morphogenesis, remodeling and pathological processes such as cancer [1]. During EMT polarized, tightly adhered epithelial cell monolayers are transformed into non-interacting motile mesenchymal cells that simultaneously degrade and synthesize extracellular matrix (ECM) components and invade into the underlying tissue space [2]. EMT is the fundamental initiator of developmental processes such as embryonic gastrulation and valvulogenesis [3] (also Kalluri J Clin Invest 2009, Thiery Cell 2009). Transforming growth factor  $\beta$  (TGF- $\beta$ ) family members are important inducers of both developmental and pathological EMT [4,5]. Decades of research has focused on identifying molecular regulators of EMT, but almost all on a single gene and in a nearly binary yes/no level of qualitative understanding. Medici and coworkers recently identified a core signaling program by which TGF- $\beta$  isoforms induce EMT across a variety of cell lines [6,7]. This program involves carefully orchestrated rounds of gene expression driven by the Smad and Snail families of transcription factors as well as other key factors such as lymphoid enhancer-binding factor 1 (LEF-1), nuclear factor of activated T-cells, cytoplasmic 1 (NFATc1), and specificity protein 1 (Sp1). Coregulators such as  $\beta$ -catenin, NF- $\kappa$ B, and the ErbB family of receptor tyrosine kinases however also participate in EMT regulation, but the degree of each's influence is difficult to ascertain in isolation [8–11]. EMT also exhibits complex temporal dynamics that are often intractable in gain/loss of function studies. Elucidating the master regulatory architecture controlling EMT therefore requires inclusion of these complex overlapping and non-binary

behaviors.

Systems biology and mathematical modeling are essential tools for understanding complex developmental programs like EMT [12]. Previous computational models of TGF- $\beta$  induced differentiation focused on single biological factors or EMT in single cells. For example, Chung *et al.*, constructed a model of TGF- $\beta$  receptor activation and Smad signaling using ordinary differential equations and mass-action kinetics. Their model suggested that a reduction of functional TGF- $\beta$  receptors in cancer cells may lead to an attenuated Smad2 signal [13]. Similarly, Vilar *et al.* suggested that specific changes in receptor trafficking patterns could lead to phenotypes that favor tumorigenesis [14]. Although these models provided insight into the role of receptor dynamics, EMT induction involves many other components, including competing second messengers and interconnected transcriptional regulatory loops. Integrating these additional scales of molecular signaling while maintaining the capacity for robust prediction requires a new and expanded computational and experimental strategy. Data-driven systems approaches [15] or logical model formulations [16] are emerging paradigms that constrain model complexity through the incorporation of training and validation data. These are interesting techniques because the data informs model structure (which can be expanded as more data becomes available). Alternatively, Bailey proposed more than a decade ago that a qualitative understanding of a complex biological system should not require complete definition of its structural and parametric content [17]. Shortly thereafter, Sethna and coworkers showed that complex model behavior is often controlled by only a few parameter combinations, a characteristic seemingly universal to multi-parameter models referred to as “sloppiness” [18]. Thus, reasonable model predictions are often possible with only limited parameter information. Taking advantage of this property, we developed sloppy techniques for parameter identification using ensembles of deterministic models [19]. Furthermore, we proposed that the sloppy behavior of biological networks may also be seen as a source of cell-to-cell [20] or even patient-to-patient heterogeneity [21]. Recently, Bayesian parameter identification techniques have also been used to explore cell-to-cell heterogeneity [22,23], where a population of cells could be viewed as a dynamic ensemble of context-specific biochemical networks [24].

In this study, we developed a family of mechanistic models describing the induction of EMT by TGF- $\beta$  isoforms in the presence and absence of vascular endothelial growth factor A (VEGF-A). We incorporated mass action kinetics within an ordinary differential equation (ODE) framework to describe the EMT interaction network containing 995 gene, protein or mRNA components interconnected through 1700 interactions. A family of model parameters was estimated using 41 molecular data sets generated in DLD1 colon carcinoma, MDCKII and A375 melanoma cells using the Pareto optimal ensemble technique (POETs) multiobjective optimization algorithm. POETs identified more than 15,000 likely TGF- $\beta$  induced EMT models, from which we selected approximately 1100 models for further analysis. Analysis of the model population suggested that both MCF10A and DLD1 cells could exhibit

phenotypic heterogeneity if treated simultaneously with TGF- $\beta$ 1/2 and VEGF-A. This heterogeneity was characterized by a significant fraction of the population being in a “hybrid state” having both high E-cadherin and high Vimentin expression. We tested these predictions using qRT-PCR and flow-cytometry studies in a variety of experimental conditions. Validation studies confirmed that upwards of 45% of the cellular population could be put into the hybrid state in the presence of both TGF- $\beta$ 1/2 and VEGF-A. Moreover, this response depended upon both activation of Sp1 by MAPK and NFATc1 transcriptional activity consistent with the predicted molecular signaling. Lastly, the hybrid populations of both DLD1 and MCF10A cells exhibited different functional behavior than those from either TGF- $\beta$  or VEGF-A treatment. The extent of ductal branch formation significantly increased with MCF10A cells in the hybrid phenotype, compared with cells treated with VEGF-A alone. Together, these results establish a predictive mechanistic model of EMT susceptibility, and reveal a novel signaling axis, which possibly regulates carcinoma progression through an EMT versus tubulogenesis response.

## Results

### The model population captured key features of TGF- $\beta$ induced EMT

The EMT model architecture, based upon curated molecular connectivity, described the expression of 80 genes following exposure to TGF- $\beta$  isoforms and VEGF-A (Fig. 1). The EMT model contained 995 molecular species interconnected by 1700 interactions. Model equations were formulated using mass-action kinetics within an ordinary differential equation (ODE) framework. ODEs and mass action kinetics are common tools to model biochemical pathways [25–27]. However, while ODE models can simulate complex intracellular behavior, they require estimates for model parameters which are often difficult to obtain. The EMT model had 1756 unknown model parameters, 1700 kinetic constants and 56 non-zero initial conditions. As expected, these parameters were not uniquely identifiable given the training data [28]. Thus, instead of identifying a single best fit (but uncertain) model, we estimated a sloppy population of models (each consistent with the training data) by simultaneously minimizing the difference between model simulations and 41 molecular data sets using the Pareto Optimal Ensemble Technique (POETs). The training data were generated in DLD1 colon carcinoma, MDCKII, and A375 melanoma cells following exposure to TGF- $\beta$  isoforms [7]. We organized these data sets into 11 objective functions which were simultaneously minimized by POETs. Additionally, we used 12 molecular data sets generated in HK-2 cells following VEGF-A exposure to train VEGF-A responsive model processes [29]. To guard against overfitting, we augmented the multiobjective optimization with leave-one-out cross validation to independently estimate both the training and prediction error for each objective. Thus, we generated 11 different model ensembles. Lastly, we compared model predictions with independent data sets not used during training (both at the molecular and model population levels) to evaluate the predictive power of the parameter ensemble. Additional details of the signaling architecture included in the model are presented in the materials and methods and the supplement.

POETs generated a population of probable signaling models which captured the multiple phases of EMT induction (Fig. 2). POETs sampled well over  $10^6$  probable models during each stage of the cross-validation, using a combination of both local and global random sampling. This sampling generated approximately 15,000 highly probable models from which we selected  $N \simeq 1100$  models for further analysis. The selected models all had the same possible molecular connectivity, but different values for model parameters and extrinsic factors such as RNA polymerase or ribosome abundance. Model selection was based upon Pareto rank, the prediction and training error across all objectives and model to model correlation (supplemental materials). The model population recapitulated key signaling events following TGF- $\beta$  exposure. We subdivided the response to TGF- $\beta$  exposure into two phases. First, TGF- $\beta 1/2$  signaling initiated a program which downregulated E-cadherin expression in a MAPK dependent manner while simultaneously upregulating TGF- $\beta 3$  expression.

Second, TGF- $\beta$ 3 secretion initiated an autocrine feedback which upregulated the expression of mesenchymal markers such as Vimentin and key upstream transcription factors such as LEF-1 in a SMAD dependent manner. Each phase involved the hierachal expression and/or post-translational modification of several key transcription factors. During the first phase, stimulation with TGF- $\beta$ 1/2 (10 a.u.) activated both the SMAD and MAPK pathways. MAPK activation resulted in the phosphorylation of the transcription factor activator protein 1 (AP-1), which in-turn upregulated the expression of Snail, a well established transcriptional repressor (Fig. 2A). Snail expression was MAPK-dependent; the MEK inhibitor U0126 blocked AP-1 activation and Snail expression following TGF- $\beta$ 1/2 exposure (Fig. 2A, Lane 3). Similar results were obtained for Slug expression, confirming initial activation through the MAPK pathway (data not shown). Overexpression of either Snail or Slug upregulated TGF- $\beta$ 3 expression (Fig. 2C) while simultaneously downregulating E-cadherin expression (Fig. 2F). During the second phase, TGF- $\beta$ 3 secretion and the subsequent autocrine signaling resulted in the upregulation of mesenchymal marker expression. The TGF- $\beta$ 3 induced gene expression program involves a complex hierarchy of transcriptional and post-translational regulatory events. Absence of E-cadherin indirectly promoted TGF- $\beta$ 3 expression through the  $\beta$ -catenin/TCF4 complex following Snail or Slug expression (Fig. 2C, Lane 2 or 3). Conversely, over-expression of E-cadherin inhibited the TGF- $\beta$ 3 autocrine production by sequestering cytosolic  $\beta$ -catenin, thereby blocking EMT (Fig. 2C, Lane 4 or 5). TGF- $\beta$ 3 signaled through the Smad pathway to regulate LEF-1 expression and downstream target EMT genes (Fig. 2G). TGF- $\beta$ 3 (10 a.u.) in combination with downstream inhibitors (DN-Smad4 and DN-LEF-1) completely inhibited Vimentin expression, while elevating E-cadherin expression (Fig. 2H,I).

The predictive power of the ensemble was tested using both cross validation and by comparing simulations with data sets not used for model training. In whole, 78% of our training objectives were statistically significant (at a 95% confidence interval) compared to a randomized parameter family ( $N = 100$ ) generated from the best-fit nominal set (starting point for the optimization). Conversely, we *predicted* approximately 60% of the training objectives, at a 95% confidence interval compared to randomized parameters. The model also captured the temporal gene expression responses of E-cadherin, pSmad2, and LEF-1 to within one-standard deviation (up to the 48 hr time-point) (Fig. 2J-L). This data was not used for model training. The high predictability can be attributed to the combination of the leave-one-out cross validation scheme, diverse objective functions, and robustness of the POETs algorithm. Taken together, the model captured the key signaling events revealed by Medici *et al.* [7] that drive the phenotypic conversion. A listing of data used for training is included in the supplement (Fig. S5 and Fig. S6).

### Identification of a novel LEF-1 regulator

During model identification, we found that consistent TGF- $\beta$  induced EMT required an additional regulatory protein. This protein, which we called

hypothetical regulator 1 (YREG1), was required to mediate between  
178  
SNAIL/SLUG transcriptional activity and the upregulation of LEF-1 expression  
179 following TGF- $\beta$ 1/2 exposure. SNAIL/SLUG are well known transcriptional  
180 repressors [30–32], although there are a few studies which suggest that at least  
181 SNAIL can also act as a transcriptional activator [33]. In the model, we assumed  
182 the expression of SNAIL/SLUG was likely regulated by AP1/SP1 [34]. Thus,  
183 upon receiving a TGF- $\beta$ 1/2 signal, the model predicted enhanced SNAIL/SLUG  
184 expression, consistent with experimental observations. TGF- $\beta$ 1/2 stimulation  
185 also induces LEF-1 expression. However, literature evidence suggested that  
186 LEF-1 expression was not strongly dependent upon AP1/SP1 activity [35]. Thus,  
187 either SNAIL/SLUG are acting as inducers (contrary to substantial biochemical  
188 evidence) or, they are repressing the expression of an intermediate repressor.  
189 Given the biochemical evidence supporting SNAIL/SLUG as repressors, we  
190 created YREG1 a hypothetical intermediate repressor whose expression is  
191 downregulated by SNAIL/SLUG. The literature data therefore suggested that  
192 YREG1 had two transcriptional targets, LEF-1 and TGF- $\beta$ 3. By adding this  
193 regulator, our simulations became consistent with training and literature data.  
194 Medici et al. suggested a similar idea where feedback between  $\beta$ -catenin and  
195 LEF-1 was likely, although this feedback had yet to be identified [7]. Low levels  
196 of YREG1 expression were used in all simulations to regulate the formation of  
197 the  $\beta$ -catenin-LEF-1 complex. To test the potency of YREG1, we conducted  
198 knockdown and over-expression simulations following the addition of TGF- $\beta$ 1/2  
199 (Fig. S8). In the absence of YREG1, most of the population failed to  
200 consistently respond to TGF- $\beta$ 1/2 exposure compared to the wild-type (Fig.  
201 S8A). Conversely, YREG1 overexpression revealed an exclusively epithelial  
202 phenotype following TGF- $\beta$ 1/2 stimulation (Fig. S8B). Overexpression of  
203 YREG1 repressed LEF-1 and TGF- $\beta$ 3 expression, thereby not allowing free  
204  $\beta$ -catenin to form the  $\beta$ -catenin-LEF-1 complex which promotes mesenchymal  
205 gene expression, or SMAD activity following from autocrine TGF- $\beta$ 3 signaling.  
206 Likewise, the abundance of the pSmad2/4-LEF-1 complex was also reduced in  
207 cells overexpression YREG1, which blocked the repression of E-cadherin. Taken  
208 together, we found that low YREG1 expression was necessary for stabilizing  
209 EMT, while elevated YREG1 levels limited the extent of EMT induction.  
210

## TGF- $\beta$ 1/2 and VEGF-A exposure promotes phenotype heterogeneity through NFATc and phosphorylated Sp1

While we captured the central tendency of many of the molecular features of EMT induction following TGF- $\beta$ 1/2 exposure, an often neglected but important emergent feature of developmental and pathological programs is population heterogeneity [36]. We (and others) previously hypothesized that deterministic model ensembles can interrogate population behavior, at least at a coarse grained level [20]. We tested this hypothesis by analyzing the response of the population of EMT models to extracellular cues and then comparing this response to flow cytometry studies. We used robustness coefficients to quantify the response of the individual members of the ensemble to TGF- $\beta$ 1/2 stimulation. We have previously used robustness coefficients to systematically quantify response of a system to structural or operational perturbations, for example gene deletions or the addition of a growth factor or hormone [19, 20, 27, 37]. Robustness coefficients quantify shifts in molecular marker abundance resulting from molecular or environmental perturbations relative to an unperturbed control state. Robustness coefficients  $\gg 1$  indicate that marker abundance increased, while robustness coefficients  $\ll 1$  indicates marker abundance decreased relative to an unperturbed control. A value of  $\sim 1$  indicates approximately no change in marker abundance following the perturbation. We calculated robustness coefficients for each member of the ensemble ( $N \simeq 1100$ ) for two downstream phenotypic markers, Vimentin (mesenchymal) and E-cadherin (epithelial) following the addition of TGF- $\beta$ 1/2 alone (Fig. 3), and VEGF-A in combination with NFATc inhibitors (Fig. 4). The absence of TGF- $\beta$ 1/2 or VEGF-A stimulation was used as the baseline for the robustness calculations.

We identified model subpopulations that exhibited different behaviors following exposure to TGF- $\beta$ 1/2 (Fig. 3A, labeled P1-P4). Analysis of the molecular signatures of these subpopulations suggested the abundance, localization and state of the Sp1, AP-1 and NFATc transcription factors controlled population heterogeneity. The behavior of the majority of models (>70%) was similar to subpopulation one (P1) or subpopulation two (P2) in Fig 3. These models showed the classically expected behavior, a switch from an epithelial to mesenchymal phenotype following TGF- $\beta$ 1/2 exposure. Models near P1 had elevated nuclear localized phosphorylated Sp1, relative to non-induced cells (and models near P2). Elevated Sp1 activity decreased E-cadherin expression through Slug-mediated inhibition, which in turn increased Vimentin expression through TGF- $\beta$ 3 autocrine signaling and the liberation of  $\beta$ -catenin. Near P2, Sp1 transcriptional activity was lower than P1, leading to only modestly increased Vimentin expression and E-cadherin repression following TGF- $\beta$ 1/2 stimulation. Near subpopulation three (P3), reduced levels of nuclear phosphorylated AP-1, Sp1, and NFAT (resulting from the loss of ERK kinase activity) were responsible for Vimentin *repression* relative to the control. However, the most biologically interesting behavior was exhibited by subpopulation four (P4). Models near P4 had elevated Sp1 and NFAT transcriptional activity, which increased *both* Vimentin and E-cadherin

expression. Analysis of these hypothetical cells suggested they had *abnormal*  
 257  
 signaling; deregulated NFAT expression and nuclear localization promoted  
 258  
 E-cadherin expression while TGF- $\beta$ 1/2 induced Sp1 action promoted Vimentin  
 259  
 expression. Analysis of the connectivity and information flow through the  
 260  
 signaling architecture suggested that Sp1 and NFAT action could be  
 261  
 manipulated *independently* by simultaneous TGF- $\beta$ 1/2 and VEGF-A  
 262  
 stimulation (Fig. S1).  
 263

To test this hypothesis, we simulated the response of the network to  
 264  
 TGF- $\beta$ 1/2 and VEGF-A treatment with and without NFATc inhibitors (Fig.  
 265  
 4). As expected, stimulation with VEGF-A (50 a.u.) maintained an epithelial  
 266  
 population (Q4-43.6%), while TGF- $\beta$ 1/2 (10 a.u.) exposure shifted the  
 267  
 population from an epithelial (Q4-5.5%) to a mesenchymal (Q1-45.6%)  
 268  
 phenotype (Fig. 4A and Fig. 4B). On the other hand, combined stimulation  
 269  
 with TGF- $\beta$ 1/2 (10 a.u.) and VEGF-A (50 a.u.) increased both E-cadherin  
 270  
 and Vimentin expression (Q2-45.3%), resulting in a hybrid phenotype with both  
 271  
 epithelial and mesenchymal characteristics (Fig. 4C). To better understand this  
 272  
 hybrid response, we quantified the simulated protein levels for E-cadherin,  
 273  
 Vimentin, phosphorylated nuclear Sp1, nuclear NFATc1,  $\alpha$ -smooth muscle actin  
 274  
 ( $\alpha$ -SMA) and Slug as a function of condition (Fig. S2A-C). Vimentin expression  
 275  
 was correlated with high levels of nuclear phosphorylated Sp1, following  
 276  
 TGF- $\beta$ 1/2 exposure. Conversely, elevated E-cadherin expression depended  
 277  
 upon the activity of NFAT transcription factors downstream of VEGF-A  
 278  
 stimulation. To further isolate the role of NFAT on this hybrid state, we  
 279  
 simulated the inhibition of NFAT transcriptional activity across all conditions  
 280  
 (all else being equal). NFAT inhibition in combination with VEGF-A treatment  
 281  
 blocked all E-cadherin positive sets (Fig. 4D). Likewise, TGF- $\beta$ 1/2 treatment  
 282  
 in combination with NFATc inhibition also resulted in the loss of E-cadherin  
 283  
 expression (Fig. 4E). Lastly, NFATc inhibition in combination with simultaneous  
 284  
 TGF- $\beta$ 1/2 and VEGF-A exposure repressed nearly all E-cadherin expression,  
 285  
 shifting nearly the entire population towards a mesenchymal phenotype (Fig.  
 286  
 4F). Taken together, high levels of nuclear localized phosphorylated Sp1  
 287  
 correlated with Vimentin expression, while NFATc transcriptional activity was  
 288  
 predicted to be critical for maintaining E-cadherin expression.  
 289

### Combined TGF- $\beta$ 2 and VEGF-A exposure drives heterogeneity 290 in MCF10A and DLD1 cells

The EMT model simulations suggested the transcriptional activity of NFATc  
 291  
 and Sp1 could be independently tuned to generate a hybrid cell population with  
 292  
 both epithelial and mesenchymal characteristics. To test this hypothesis, we  
 293  
 exposed either quiescent epithelial (MCF10, (Fig. 5)) or transformed epithelial  
 294  
 cells (DLD1, (Fig. S3)) to combinations of TGF- $\beta$ 1/2 and/or VEGF-A. As  
 295  
 expected, treatment with TGF- $\beta$ 1/2 (10ng/ml) increased Slug and Vimentin  
 296  
 expression, while repressing E-cadherin expression both at the transcript and  
 297  
 protein levels in MCF10A (Fig. 5A-B) and DLD1 cells (Fig. S4C, Fig S3 D,E).  
 298  
 Both MCF10A (Fig. 5C) and DLD1 cells (Fig. S3E,G) transitioned from  
 299  
 quiescent cobblestone morphology to spread spindle shapes, consistent with  
 300  
 301

EMT. As predicted, we found increased nuclear localization of phosphorylated Sp1 following TGF- $\beta$ 1/2 stimulation in both MCF10A (Fig. 5B,C) and DLD1 cells (Fig. S3E,F). Consistent with model predictions, VEGF-A (50ng/ml) treatment increased the abundance of NFATc1 and E-cadherin at both the transcript and protein level in both MCF10A (Fig. 5A) and DLD1 (Fig. S3A) cells. We also found that NFATc1 nuclear localization significantly increased in both MCF10 and DLD1 treated with VEGF-A independently of the abundance of nuclear localized phosphorylated Sp1 levels (Fig. 5B,C Fig.S3C,E ). Interestingly, combining VEGF-A (50ng/ml) with TGF- $\beta$ 1/2 (10ng/ml) resulted in significantly elevated expression of both E-cadherin and Vimentin at the transcript and protein levels in both MCF10A and DLD1 cells (Fig 5A,B; Fig S3D,E; Fig S4C). NFATc1 expression increased, while Sp1 expression was similar to the TGF- $\beta$ 1/2 case alone (Fig. 5A-B, Fig S3D,E; Fig S4C)), supporting their independent regulation. The expression of Slug, and Vimentin significantly increased, while E-cadherin levels were increased in MCF10A cells (Fig 5A) and maintained at control levels in DLD1 cells (Fig. S3D). As further predicted, nuclear co-localization of both NFATc1 and phosphorylated Sp1 were apparent in MCF10A and DLD1 cells treated with both ligands (Fig. 5B,C Fig S3E,F). Taken together, combined VEGF-A and TGF- $\beta$ 1/2 treatment elicited a hybrid phenotype expressing both mesenchymal and epithelial characteristics in both MCF10A and DLD1 cells. This phenotype was driven by the transcriptional activity of two key transcription factors, Sp1 and NFATc, which could be modulated independently by TGF- $\beta$ 1/2 and VEGF-A exposure.

Our robustness analysis predicted that NFATc transcriptional activity was critical to maintaining E-cadherin expression in the presence of both VEGF-A and TGF- $\beta$ 1/2. We experimentally tested this hypothesis by exposing both MCF10A (Fig. 5E,F) and DLD1 cells (Fig. S4) to combinations of VEGF-A and TGF- $\beta$ 1/2 in the presence or absence of VIVIT, a soluble peptide inhibitor of NFATc transcriptional activity [38]. Treatment with VEGF-A (50ng/ml) and VIVIT (10 $\mu$ M) in MCF10A cells significantly reduced E-cadherin expression compared to VEGF-A alone (Fig 5D,E). Co-treatment with VIVIT and TGF- $\beta$ 1/2 did not enhance EMT capacity of MCF10A cells above that of TGF- $\beta$ 1/2 alone (Fig 5A,B,E). Likewise, VIVIT in combination with both TGF- $\beta$ 1/2 and VEGF-A resulted in a loss of E-cadherin gene and protein expression, while Slug and Vimentin levels remained increased (Fig. 5D,E ). Quantitative flow cytometry confirmed these results in both MCF10A (Fig. 5F) and DLD1 cells (Fig. S4C). Both epithelial cell lines initially had high levels of E-cadherin expression, and low vimentin abundance (Q1-99.5%), but both MCF10A and DLD1 cells shifted from an epithelial to mesenchymal phenotype (Q1-33.4%, Q4-42.8%) following TGF- $\beta$ 1/2 exposure. As expected, NFATc nuclear localization was repressed with VIVIT treatment regardless of ligand stimulation, while the abundance of nuclear phosphorylated Sp1 increased for both TGF- $\beta$ 1/2 and TGF- $\beta$ 1/2 + VIVIT conditions (Fig. 5D,E). Combined TGF- $\beta$ 1/2 and VEGF-A increased both Vimentin and E-cadherin expression (Q1-42.1%, Q2-52.3%) compared to TGF- $\beta$ 1/2 alone. Together, these results demonstrate that NFATc and phosphorylated Sp1 are critical for regulating

E-cadherin and Vimentin expression during phenotype heterogeneity in MCF10A and DLD1. 348  
349

**Ductal branching during acini formation is dependent upon phenotype heterogeneity in MCF10A and DLD1 cells** 350  
351

We finally employed established three-dimensional (3D) *in vitro* models of invasion, migration, compaction, and tubulogenesis [39] to determine the functional consequences of the hybrid phenotype (Fig. 6). MCF10A and DLD1 cells were aggregated via hanging drop, placed on the surface of a collagen gel, and cultured for 72 hrs under various biochemical treatments. TGF- $\beta$ 1/2 stimulation significantly enhanced cell matrix invasion and matrix compaction, while in contrast VEGF-A stimulation promoted surface migration but no invasion or compaction (Fig. 6B-D). Interestingly, combined TGF- $\beta$ 1/2 and VEGF-A stimulation significantly increased cell migration potential above that of VEGF-A alone while maintaining 3D matrix compaction, though with decreased magnitude compared to TGF- $\beta$ 1/2 alone. Inhibition of NFATc transcriptional activity by VIVIT decreased migration following treatment with VEGF-A alone (Fig. 6B). Co-treatment of VIVIT significantly decreased migration, while complementarily increasing invasion and compaction, when MCF10A cells were stimulated with both VEGF-A and TGF- $\beta$ 1/2 (Fig. 6B-D). The responses of DLD1 cells followed a similar trend to MCF10A, although the magnitudes of migration, invasion, and compaction were less. Cell circularity within 3D gels strongly and negatively correlated with both invasion and compaction regardless of treatment (Fig. 6E). Circularity refers to the morphology of the cells. In general, a quiescent epithelial cells assumes a circular morphology in culture, while an active mesenchymal cell is highly elongated. The circularity index, a common means of quantifying cell morphology, relates cell area to perimeter. A perfect circle has a circularity index equal to 1.0, while a straight line has a circularity index equal to 0.0, see Butcher et al. [40]. TGF- $\beta$ 1/2 treatment alone resulted in irregular and spindle shaped morphology, while VEGF-A exposure promoted round quiescent cells (Fig. 6A). Combined VEGF-A and TGF- $\beta$ 1/2 promoted morphology between these extremes. VIVIT mediated NFATc inhibition significantly reduced the circularity index, similar to TGF- $\beta$ 1/2 treatment (Fig. 6F). VEGF-A treatment also induced the formation of tubular structures (acini), but the number of tubular branches relative to total acini was significantly increased upon combined TGF- $\beta$ 1/2 and VEGF-A. No tubular structures were identified within the DLD1 constructs during the 7 day tubulogenesis endpoints, supporting that MCF10A and DLD1 cells have some cell-type specific EMT sensitivity despite their underlying competency for acquiring a heterogeneous phenotype. This suggests that initial EMT sensitivity of a cell influences downstream functional response from TGF- $\beta$  and VEGFA stimulation. Together, these results establish that VEGF-A and TGF- $\beta$ 1/2 ligand concentrations potentiate between acini and ductal branch formation in 3D culture, and are dependent upon NFATc activity.

## Discussion

391

In this study, we developed a family of mechanistic models describing the induction of EMT by TGF- $\beta$  isoforms in the presence and absence of VEGF-A. The signaling architecture encoded in the model, which contained 995 molecular species interconnected by 1700 interactions, described the expression of 80 genes in response to growth factor stimulation. This simulation incorporates an unprecedented level of detail compared to previous models, but as a consequence created a large number of unknown model parameters. Because these parameters could not be estimated uniquely apriori, we estimated an ensemble of likely parameters using the POETs multiobjective optimization framework. The model population was trained and cross-validated to prescribe biological significance using 41 data sets generated in DLD1 colon carcinoma, MDCKII, and A375 melanoma cell lines [7]. POETs generated  $> 15,000$  probable parameter sets using this data, from which we selected  $N \simeq 1100$  for subsequent analysis. Analysis of this population predicted possible phenotypic modes (and their associated signaling) that cells could exhibit when stimulated with TGF- $\beta$  and/or VEGF-A. The most novel hypothesis generated from the analysis was that cells could operate in a hybrid state defined by both epithelial and mesenchymal traits when stimulated simultaneously with TGF- $\beta$  and VEGF-A. We tested this hypothesis in MCF10A and DLD1 cells stimulated with combinations of TGF- $\beta$  and VEGF-A. As expected, in the presence of TGF- $\beta$  or VEGF-A alone, MCF10A and DLD1 cells were either mesenchymal or epithelial, respectively. However, with both TGF- $\beta$  and VEGF-A, MCF10A and DLD1 cells exhibited a hybrid phenotype, having both epithelial and mesenchymal characteristics. Furthermore, we found that functional traits such as tubulogenesis and ductal branching were different for cells in this hybrid phenotype. Together, this study established a predictive model of EMT induction, determined that deterministic model ensembles could predict population heterogeneity, and proved the existence of a unique hybrid phenotype resulting from the simultaneous integration of extracellular growth factor signals.

Cells routinely process a multitude of signals simultaneously, especially when coordinating developmental or pathological programs. For example, oncogenic cells integrate both mechanical and chemical cues in their local microenvironment during tumorigenesis, including cytokines VEGF and TGF- $\beta$  [41]. VEGF-A mediates pathological angiogenic remodeling of tumors [42], while TGF- $\beta$  can elicit both protective and oncogenic responses [43,44]. While much research has tested signaling pathways individually, far less is understood about combinatorial stimulation, such as with both VEGF-A and TGF- $\beta$ . Recent *in vitro* and *in vivo* evidence has suggested that epithelial cells can exhibit heterogeneous phenotypes in addition to classically defined epithelial or mesenchymal states [45,46]. For example, expression profiling in human epithelial cancer cell lines demonstrated a spectrum of phenotypes, including some that expressed both E-cadherin and Vimentin simultaneously [47,48]. Zajchowski *et al.*, speculated that these expression profiles were somehow important for maintaining epithelial properties,

while simultaneously allowing other functional behavior such as proliferation and migration [49]. Whether and how heterogeneous phenotypes arise and participate in cancer progression, as well as their response to pharmacological inhibition are fundamental questions that should receive increased attention. In this study, we determined that a hybrid phenotype could be obtained through combined treatment with VEGF-A and TGF- $\beta$ , both common factors localized in the tumor microenvironment. Furthermore, our systematic simulation-experimentation strategy identified that the transcriptional activity of Sp1 and NFATc were the critical factors controlling this phenotypic heterogeneity. Several studies have highlighted the importance of NFATc as a key transcription factor involved in cell growth, survival, invasion, angiogenesis and cancer [50]. For example, proliferation and anchorage-independent growth of pancreatic tumor cells is dependent on calcineurin and NFATc1 activity, consistent with the high levels of nuclear NFATc1 found in pancreatic tumors [51]. Likewise, our results found that VEGF-A was a potent inducer of NFATc1 expression, which may be required for epithelial cell migration and tubulogenesis. Although specific NFATc isoforms were not distinguished in the model, our simulations suggested that NFATc transcriptional activity was capable of maintaining epithelial traits, even during TGF- $\beta$  induced EMT. Experimentally, we found that E-cadherin expression was dependent upon NFATc dephosphorylation in response to simultaneous VEGF-A and TGF- $\beta$ 1/2 treatment. Thus, these results support the hypothesis that NFATc activity plays a critical role in maintaining cell-cell contacts, even during partial EMT.

Epithelial cells reproduce tissue-like organization when grown in a three-dimensional extracellular matrix (ECM) environment, and therefore are an attractive model to study morphogenic mechanisms. It is well established that MCF10A cells form structures that closely resemble acini (multi-lobed cluster of cells) in three-dimensional *in vitro* cultures [52]. It has been postulated that a cellular response reminiscent of partial EMT underlies this process, stimulating further branching and formation of acini [53]. Normally well controlled process such as tubulogenesis can be co-opted by cancer cells to break away from a primary lesion and invade through the surrounding stroma [54]. However, by retaining a transient hybrid EMT-like state, clusters of these tube-forming tumor cells can reform at a high rate after invasion, possibly explaining why invasive human carcinomas frequently appear to be cellular collections with varying degrees of gland-like differentiation [55]. In this study, we showed that our predicted hybrid phenotype generated by simultaneous treatment of epithelial cells with VEGF-A and TGF- $\beta$  possessed altered migration and invasion, which enhanced tubular branching. A salient feature of this behavior, however, was the retention of cell-cell contacts that allowed cells to migrate without completely dissociating from their neighbors. Thus, our results support a mechanism in which hybrid cells can maintain some functional characteristics of epithelial cells such as cell-cell adhesion, which are normally lost in a fully differentiated mesenchymal state. The tumor microenvironment contains many soluble signals simultaneously, including VEGF and TGF- $\beta$ . Thus, it is likely that some cancerous epithelial cells could exhibit hybrid EMT phenotypic states.

This may explain why fibroblastoid morphology, a classical feature of EMT, is  
482  
not commonly observed in human carcinomas [55]. This study focused on the  
483  
combinatorial effects of two very different ligand families present together in the  
484  
tumor environment. Additional modeling studies are required to unravel the  
485  
global response of epithelial cells to the full spectrum of chemical, substrate, and  
486  
mechanical cues. The simulation strategy presented here is readily adaptable to  
487  
larger species sets, with the major advantage that experimentally testable  
488  
hypotheses can be generated regarding how signals get integrated to produce  
489  
global cellular response. Furthermore, by simulating multiple ensembles of  
490  
parameter sets, subpopulations across a constellation of phenotypes can be  
491  
created and mined for common and/or divergent signaling characteristics. This  
492  
is a significant advantage over forced convergence to a single unique solution and  
493  
thereby generating a potentially non-physiological homogeneous population.  
494

The deterministic population of EMT models predicted heterogeneous  
495  
behavior that was qualitatively consistent with experimental studies. There is a  
496  
diversity of algorithmic approaches to estimate model parameters [56], as well as  
497  
many strategies to integrate model identification with experimental  
498  
design [57, 58]. However, despite these advances, the identification of models  
499  
describing intracellular network behavior remains challenging. There are  
500  
different schools of thought to deal with this challenge. One school has focused  
501  
on model reduction. Data-driven approaches [15], boolean [59] or other logical  
502  
model formulations [16, 60] are emerging paradigms that constrain model  
503  
complexity by the availability of the training and validation data. Other  
504  
techniques such as constraints based modeling, which is commonly used to  
505  
model metabolic networks, have also been applied to model transcriptional  
506  
networks, although primarily in lower eukaryotes and prokaryotes [61]. These  
507  
techniques (and many others, see review [62]) are certainly exciting, with many  
508  
interesting properties. However, we used the traditional approach of mass action  
509  
kinetics within an ordinary differential equation framework. The identification  
510  
problem for the EMT model was massively underdetermined. This is not  
511  
uncommon for differential equation models, especially those that are highly  
512  
mechanistic. Of course, we could have discarded mechanism or reduced the  
513  
model scope to decrease the complexity of the identification problem. However,  
514  
a central criticism leveled by biologists is that model simplification is often done  
515  
at the cost of biological reality, or done for reasons of computational  
516  
expediency [63]. To avoid this criticism, we systematically identified an  
517  
ensemble of likely models each consistent with the training data, instead of a  
518  
single but uncertain best fit model. Previously, we (and others) have suggested  
519  
that deterministic ensembles could model heterogeneous populations in  
520  
situations where stochastic computation was not feasible [20]. Population  
521  
heterogeneity using deterministic model families has previously been explored  
522  
for bacterial growth in batch cultures [64]. In that case, distributions were  
523  
generated because the model parameters varied over the ensemble, i.e., extrinsic  
524  
noise led to population heterogeneity. In this study, parameters controlling  
525  
physical interactions such as disassociation rates or the rate of assembly or  
526  
degradation of macromolecular machinery such as ribosomes were widely  
527

distributed over the ensemble. Population heterogeneity can also arise from 528  
intrinsic thermal fluctuations, which are not captured by a deterministic 529  
population of models [65]. Thus, deterministic ensembles, provide a 530  
coarse-grained or extrinsic-only ability to simulate population diversity. Despite 531  
this limitation, our prediction of phenotypic heterogeneity (and the underlying 532  
signaling events responsible for the heterogeneity) was consistent with 533  
experimental observations. This suggested that deterministic ensembles could 534  
simulate disease or developmental processes in which heterogeneity plays an 535  
important role, without having to resort to stochastic simulation. 536

A common criticism of ODE modeling has been the poorly characterized 537  
effect of structural and parametric uncertainty. In this study, parametric 538  
uncertainty was addressed by developing an ensemble of probable models 539  
instead of a single best-fit but uncertain model using multiobjective 540  
optimization. While computationally complex, multiobjective optimization is an 541  
important tool to address qualitative conflicts in training data that arise from 542  
experimental error or cell line artifacts [66]. On the other hand, structural 543  
uncertainty is defined as uncertainty in the biological connectivity. The EMT 544  
model connectivity was assembled from an extensive literature review. However, 545  
several potentially important signaling mechanisms were not included. First, we 546  
identified a potential gap in biological knowledge surrounding the regulation of 547  
LEF-1 expression, that was filled by the addition of the hypothetical YREG1 548  
transcriptional repressor. The LEF-1 transcription factor is expressed in tissues 549  
that undergo EMT during embryogenesis [67, 68], and has been suggested to 550  
promote an invasive phenotype in cancer cells [8, 69]. Low levels of YREG1 were 551  
important for stabilizing the interaction between LEF-1 and  $\beta$ -catenin, while 552  
elevated levels inhibited EMT by downregulating LEF-1 transcriptional activity. 553  
Recent evidence has established a complex role of Amino terminal Enhancer of 554  
Split (AES) and Groucho/TLE on suppressing LEF-1 activity. AES opposes 555  
LEF-1 transcriptional activation while Groucho/TLE binds with LEF-1 for a 556  
histone deacetylase repression. In addition,  $\beta$ -catenin directly displaces 557  
Groucho/TLE repressors from TCF/LEF-1 in Wnt-mediated transcription 558  
activation [70, 71]. Our model agrees with this newly discovered feedback system, 559  
as YREG1 regulates LEF-1 activity leading to EMT stabilization. Secondly, we 560  
should revisit the role of GSK-3 $\beta$ . GSK-3 $\beta$  is an important regulator which 561  
controls the abundance of both Snail and  $\beta$ -catenin through the 562  
ubiquitin-proteasome pathway [72, 73]. Specifically to our model, expression of 563  
Snail increases through 72 hrs. In contrast, experimental data has shown that 564  
activity of Snail peaks at 24 hrs which may be controlled by the GSK-3 $\beta$  565  
complex [6]. Recent evidence has also suggested an essential role of NF- $\kappa$ B in 566  
epithelial transformation. NF- $\kappa$ B may influence Snail expression through the 567  
AKT pathway and directly stabilize Snail activity [74]. This is particularly 568  
important for integrating inflammation pathways, such as interleukin-6 (IL-6) 569  
and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), which have been linked to EMT in 570  
pathological conditions [75]. Other pathways such as Notch have also been 571  
shown to act synergistically with TGF- $\beta$  to express Slug in the developing 572  
embryo [76]. Lastly, while we have modeled classical protein signaling, we have 573

not considered the role of regulatory RNAs on EMT. There is growing evidence  
that microRNAs (miRNAs) play a strong role in EMT, where several miRNAs,  
for example miR-21 and miR-31 are strongly associated with TGF- $\beta$   
exposure [77]. Addressing missing structural components like these, could  
generate more insight into TGF- $\beta$  signaling and its role in phenotypic  
transformation.

574

575

576

577

578

579

## Materials and Methods

The simulation code and parameter ensemble used in this study can be downloaded from GitHub ([https://github.com/jeffreyvarner/TGFb-VEGFA-Model\\_v1.git](https://github.com/jeffreyvarner/TGFb-VEGFA-Model_v1.git)).

### Signaling network connectivity

The EMT model described the gene expression program resulting from TGF- $\beta$  and VEGF-A signaling in a prototypical epithelial cell. The TGF- $\beta$ -EMT network contained 995 nodes (proteins, mRNA or genes) interconnected by 1700 interactions. The network connectivity was curated from more than 40 primary literature sources in combination with on-line databases [78, 79]. The model interactome was not specific to a single epithelial cell line. Rather, we assembled canonical pathways involved in TGF- $\beta$  and VEGF-A signaling, defaulting to human connectivity when possible. Using a canonical architecture allowed us to explore general features of TGF- $\beta$  induced EMT without cell line specific artifacts. On the other hand, because of the canonical architecture, we evaluated the simulation conclusions in several cell lines to test the generality of our conclusions.

Our signaling network reconstruction was based on Medici et al. who identified the pathways through which MDCKII, DLD1 colon carcinoma, and A375 melanoma cells transition towards a mesenchymal phenotype [7]. Sequential activation of MAPK and Smad pathways were initiated upon addition of TGF $\beta$ 1/2. Briefly, TGF $\beta$ 2 signals through the RAS-RAF-MEK-ERK pathway to up-regulate Snail and Slug expression [6]. Snail, a known repressor of junctional proteins, inhibits the expression of E-cadherin [69]. This initial repression of E-cadherin leads to a release of  $\beta$ -catenin from the cell membrane. Cytosolic  $\beta$ -catenin can then translocate to the nucleus and form transcriptional complexes with TCF-4 to drive TGF $\beta$ 3 expression [7]. TGF $\beta$ 3 signals to the cells interior by binding to type II receptors, which form heterodimers with type I receptors (ALK5) [80]. This activates the receptor serine/threonine kinase activity to phosphorylate and activate the receptor Smads 2/3 [81]. Phosphorylated Smads 2/3 (pSmad2/3) form heterodimers with partner Smad4 and translocate to the nucleus. pSmads complexes up-regulate other transcription factors, such as LEF-1. The pSmad2/4/LEF-1 has been shown to directly repress the E-cadherin gene [82]. LEF-1 also binds with  $\beta$ -catenin to upregulate mesenchymal proteins such as fibronectin [83]. The EMT gene expression program was initiated by the binding of TGF- $\beta$  isoforms to TGF- $\beta$  surface receptors. Binding of extracellular TGF- $\beta$ 1/2 with TGF- $\beta$  surface receptors I/II (TGF- $\beta$ R-I/II) initiates the assembly of adapter complexes which starts the downstream signaling program. In the model, TGF- $\beta$ 1/2 binds TGF- $\beta$ R-I/II followed by the recruitment of activin receptor-like kinase 1 (ALK1) and TGF- $\beta$  surface receptor III (TGF- $\beta$ R-III) to form the activated receptor complex [80]. Alternatively, we also included activin receptor-like kinase 5 (ALK5) recruitment in combination with Endoglin and TGF- $\beta$ R-III as a second (redundant) activated receptor

complex [84]. Complex assembly activates the serine/threonine kinase activity on the receptor, leading to the recruitment and phosphorylation of Smad partners [81]. Phosphorylated Smads2/3 (pSmad2/3) form heterodimers with partner Smad4 and then translocate to the nucleus where they act as both transcriptional activators and repressors. Nuclear pSmad2/3-Smad4 form transcriptional complexes with several genes in the model including lymphoid enhancer-binding factor 1 (*LEF-1*), Nuclear factor of activated T-cells, cytoplasmic 1 (*NFACT1*), and Specificity Protein 1 (*SP1*). On the other hand, nuclear pSmad2/3-Smad4 represses (in combination with the LEF-1 protein) the expression of E-cadherin (*Cdh1*) [82] and Cadherin 5, type 2 (VE-Cadherin encoded by *Cdh5*). Repression of E-cadherin expression is the central event in the transition from an epithelial to a mesenchymal phenotype [69]. However, this transition is not solely driven by transcriptional events. At the protein level, the repression of E-cadherin leads to a release of  $\beta$ -catenin from cell membrane. Cytosolic  $\beta$ -catenin then translocates to the nucleus and forms transcriptionally-active complexes with immunoglobulin transcription factor 2 (TCF-4) to drive TGF- $\beta$ 3 expression [7]. Simultaneously, ERK1/2-mediated phosphorylation of the AP1 and Sp1 transcription factors can also regulate transcriptional complexes involving NFAT, Slug, and Smads. Lastly, canonical pathways for processing extracellular VEGF-A, BMP and Wnt signals, in addition to the PI3K pathway were also included in the model. Additional information about the interactions included in the model, along with the Systems Biology Markup Language (SBML) file encoding these interactions are included in the supplemental materials.

## Formulation, solution and analysis of the EMT model equations

EMT was modeled using mass-action kinetics within an ordinary differential equation (ODE) framework:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \cdot \mathbf{r}(\mathbf{x}, \mathbf{k}) \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (1)$$

The quantity  $\mathbf{x}$  denotes the vector describing the abundance of protein, mRNA, and other species in the model ( $995 \times 1$ ). The stoichiometric matrix  $\mathbf{S}$  encodes the signaling architecture considered in the model ( $995 \times 1700$ ). Each row of  $\mathbf{S}$  describes a signaling component while each column describes a particular interaction. The  $(i, j)$  element of  $\mathbf{S}$ , denoted by  $\sigma_{ij}$ , describes how species  $i$  is involved with interaction  $j$ . If  $\sigma_{ij} > 0$ , species  $i$  is produced by interaction  $j$ . Conversely, If  $\sigma_{ij} < 0$ , then species  $i$  is consumed in interaction  $j$ . Lastly, if  $\sigma_{ij} = 0$ , then species  $i$  is not involved in interaction  $j$ . The term  $\mathbf{r}(\mathbf{x}, \mathbf{k})$  denotes the vector of interactions rates ( $1700 \times 1$ ). We modeled each network interaction (gene expression, translation and biochemical transformations) using elementary rate laws where all reversible interactions were split into two irreversible steps (supplemental materials). Thus, the rate expression for interaction  $q$  was given by:

$$r_q(\mathbf{x}, k_q) = k_q \prod_{j \in \{\mathbf{R}_q\}} x_j^{-\sigma_{jq}} \quad (2)$$

The set  $\{\mathbf{R}_q\}$  denotes reactants for reaction  $q$ , while  $\sigma_{jq}$  denotes the stoichiometric coefficient (element of the matrix  $\mathbf{S}$ ) governing species  $j$  in reaction  $q$ . The quantity  $k_q$  denotes the rate constant (unknown) governing reaction  $q$ . Model equations were generated in the C-programming language using the UNIVERSAL code generator, starting from an text-based input file (supplemental materials). UNIVERSAL, an open source Objective-C/Java code generator, is available as a Google Code project (<http://code.google.com/p/universal-code-generator/>). Model equations were solved using the CVODE solver in the SUNDIALS library [85] on an Apple workstation (Apple, Cupertino, CA) as previously described [27].

#### **Estimation of model parameters using multiobjective optimization.**

The EMT model had 1756 unknown parameters (1700 kinetic constants and 56 non-zero initial conditions) which were not uniquely identifiable given the training data. Instead, we estimated a population of likely models (each consistent with the training data) using 41 data sets generated in DLD1 colon carcinoma, MDCKII, and A375 melanoma cells taken from Medici *et al.* [7]. We used the Pareto Optimal Ensemble Technique (POETs) multiobjective optimization framework in combination with leave-one-out cross-validation to estimate an ensemble of model parameters [19]. Cross-validation was used to calculate both training and prediction error during the parameter estimation procedure [86]. The 41 intracellular protein and mRNA data-sets used for identification were organized into 11 objective functions. These 11 objective functions were then partitioned, where each partition contained ten training objectives and one validation objective. The training and validation data were Western blots. Thus, all model simulations were in arbitrary units. However, POETs does allow a soft constraint on the order of magnitude of the model concentration scale. In this study, we assumed the natural model concentration scale was pmol/L. We did not place a lower bound on model states. However, based on the pmol/L natural scale, we treated all values less than  $10^{-3}$  as zero (or no expression).

#### **Robustness coefficients.**

Robustness coefficients were calculated as shown previously [20, 27]. Robustness coefficients denoted by  $\alpha(i, j, t_o, t_f)$ :

$$\alpha(i, j, t_o, t_f) = \left( \int_{t_o}^{t_f} x_i(t) dt \right)^{-1} \left( \int_{t_o}^{t_f} x_i^{(j)}(t) dt \right) \quad (3)$$

quantify the response of a marker to a structural or operational perturbation to the network architecture. Here  $t_o$  and  $t_f$  denote the initial and final simulation time respectively, while  $i$  and  $j$  denote the indices for the marker and the perturbation respectively. A value of  $\alpha(i, j, t_o, t_f) > 1$ , indicates increased marker abundance, while  $\alpha(i, j, t_o, t_f) < 1$  indicates decreased marker abundance following perturbation  $j$ . If  $\alpha(i, j, t_o, t_f) \sim 1$  the  $j$ th perturbation

does not influence the abundance of marker  $i$ . Robustness coefficients were  
703  
calculated for each member of the ensemble ( $N \simeq 1100$ ).  
704

### Cell culture and experimental interrogation

DLD1 colon carcinoma, MCF10A, and HUVEC were acquired from the  
705  
American Tissue Culture Collection (Manassas, VA). Cells were grown in  
706  
culture with RPMI 1640 medium with 10% fetal bovine serum and 1%  
707  
penicillin/streptomycin for DLD1, EBM-2 supplemented with EGM-2, 5% fetal  
708  
bovine serum, and 1% penicillin/streptomycin for HUVEC, or MGEM 2  
709  
supplemented with insulin, bovine pituitary extract, cholera toxin, hEGF,  
710  
hydrocortisone, 5% horse serum, and 1% penicillin/streptomycin for MCF10A.  
711  
Cells were serum starved for 24 hours and removed from all experimental  
712  
conditions. Recombinant VEGFA165 was also removed from culture medium  
713  
prior to experimentation. Recombinant human TGF- $\beta$ 2 (R & D Systems,  
714  
Minneapolis, MN) was added to the culture medium at a concentration of 10  
715  
ng/ml and recombinant VEGFA165 at a concentration of (5ng/ml, 50ng/ml) for  
716  
all relative experiments. NFAT inhibitor (VIVIT peptide) (EMDBiosciences,  
717  
Darmstadt, Germany), was added to the culture medium at a concentration of  
718  
 $10\mu\text{M}$  for all relative experiments. Cells were passaged 1:3 or 1:4 every 3-6 d and  
719  
used between passages 4 and 8.  
720

### RT-PCR

RNA extractions were performed using a Qiagen total RNA purification kit  
722  
(Qiagen, Valencia, CA) and RNA was reverse transcribed to cDNA using the  
723  
SuperScript III RT-PCR kit with oligo(dT) primer (Invitrogen). Sufficient  
724  
quality RNA was determined by an absorbance ratio A260/A280 of 1.8-2.1,  
725  
while the quantity of RNA was determined by measuring the absorbance at  
726  
260nm (A260). Real-time PCR experiments were conducted using the SYBR  
727  
Green PCR system (Biorad, Hercules, CA) on a Biorad CFX96 cycler, with 40  
728  
cycles per sample. Cycling temperatures were as follows: denaturing, 95C;  
729  
annealing, 60C; and extension, 70C. Primers were designed to detect GAPDH,  
730  
E-cadherin, vimentin, Slug, Sp1, and NFATc1 in cDNA clones: Sp1 (F-TTG  
731  
AAA AAG GAG TTG GTG GC, R-TGC TGG TTC TGT AAG TTG GG,  
732  
Accession NG030361.1), NFATc1 (F-GCA TCA CAG GGA AGA CCG TGT C,  
733  
R-GAA GTT CAA TGT CGG AGT TTC TGA G, Accession NG029226.1).  
734  
GAPDH, E-cadherin, vimentin, and Slug primers were taken from previously  
735  
published literature [7].  
736

### Antibody Staining

Samples were fixed in 4% PFA overnight at 4C. Samples were then washed for  
738  
15 minutes on a rocker 3 times with PBS, permeabilized with 0.2% Triton-X 100  
739  
(VWR International, Radnor, PA) for 10 minutes, and washed another 3 times  
740  
with PBS. Samples were incubated overnight at 4C in a 1% BSA (Rockland  
741  
Immunochemicals, Inc., Gilbertsville, PA) blocking solution followed by another  
742  
743

4C overnight incubation with either rabbit anti-human E-cadherin 1:100  
 (Abcam, ab53033), mouse anti-human phospho-Sp1 1:100 (Abcam, ab37707),  
 mouse anti-human vimentin 1:100 (Invitrogen, V9), and rabbit anti-human  
 NFATc1 (Santa Cruz, sc-7294) 1:100. After 3 washes for 15 minutes with PBS,  
 samples were exposed to Alexa Fluor 488 or 568 conjugated (Invitrogen), species  
 specific secondary antibodies at 1:100 in 1% BSA for 2 hours at room  
 temperature. Three more washes with PBS for 15 minutes were followed by  
 incubation with either DRAQ5 far red nuclear stain (Enzo Life Sciences,  
 Plymouth Meeting, PA) at 1:1000.

## FACS

Flow cytometry for E-cadherin 1:100 (Abcam) and vimentin 1:100 expressing  
 cells was performed. Briefly, cells were trypsinized, fixed with 4% PFA for 10  
 min and then preserved in 50% methanol/PBS. Cells were kept in the -20C until  
 antibody staining was preformed. Samples were divided into multiple aliquots in  
 order to stain the proteins separately and compensate for secondary antibody  
 non-specific binding. Cells were incubated for 24 hrs at 4 C in primary antibody  
 diluted in either PBS (extracellular) or 0.2% saponin-PBS (intracellular). Cells  
 were then washed 3 times with PBS and incubated with appropriate secondary  
 antibodies and imaged using a Coulter Epics XL-MCL Flow Cytometer  
 (Coulter). All samples were compensated using appropriate background  
 subtraction and all samples were normalized using 7500 cells per flow condition.

## Three-Dimensional Culture and Tubulogenesis Assays

For invasion/migration assays, cells were resuspended in culture media, and  
 allowed to aggregate overnight in hanging drop culture (20 $\mu$ L; 20,000 cells). The  
 spherical aggregates were placed on the surface of neutralized type I collagen  
 hydrogels (1.5mg/mL) and allowed to adhere for 2 hrs before adding treatments.  
 Cultures were maintained for 72 hrs, after which they were fixed in 4% PFA and  
 slowly rehydrated using PBS. For compaction assays, cells were pelleted via  
 centrifugation and resuspended within a neutralized collagen hydrogel  
 (1.5mg/mL) solution at a density of 400,000 cells/mL. 250 $\mu$ L of gel was  
 inoculated into culture wells, which solidified after 60min. Treatments were then  
 added within 800 $\mu$ L of the culture medium without serum. Gels were liberated  
 from the surfaces of the culture wells the next day and cultured free floating for  
 an additional 3-7 days, exchanging serum free media with appropriate factors  
 every 48 hrs.

Tubulogenesis was defined as a typical nonmalignant acini structure. This  
 includes a polarized epithelial cell, hollow lumen, and the basal sides of the cell  
 are surrounded by ECM proteins (Fig. 6A, Controls or VEGF treated).  
 Previous work has shown that change in the morphological characteristics of  
 nontumorigenic MCF10A epithelial acini occur over time and exploiting them to  
 growth in 3D culture can be quantified. For example, using image segmentation,  
 Chang et al. [87] examined the elongation of the MCF10A acini at 6, 12, and 96  
 hours after a particular treatment. Polizzotti et al. [88] also suggested a

computational method to quantify acini structure based on morphological characteristics in nonmalignant, noninvasive, and invasive conditions. Adapted from these approaches, we first fluorescently labeled our cultures and captured the acini structures by 3D confocal microscopy. Next individual acini structures in the images were segmented by imageJ and labeled. We then extracted the number of ductal branches. Ductal branching was defined as any elongated cell cluster extending away from the total acini structure, which was manually segmented and counted using ImageJ. A total of 5 images for each condition were used, and approximately 12 acini were analyzed in each image. Total branching was normalized to the amount of acini present, and provides an overall general assessment to the extent of acini remodeling.

787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797

## Statistics

798  
799  
800  
801  
802  
803  
804

Results are expressed as mean  $\pm$  standard error,  $n \geq 6$ . Data was analyzed with the GraphPad Prism version 4.00 for Windows (GraphPad Software, San Diego, CA) and SAS (Statistical Analysis Software, Cary, NC). A one-way ANOVA with Tukey's post hoc was used to compare differences between means and data was transformed when necessary to obtain equal sample variances. Differences between means were considered significant at  $p < 0.05$ .

## Acknowledgments

805  
806  
807  
808  
809  
810  
811  
812

The project described was supported by Award Number #U54CA143876 from the National Cancer Institute to JDV, National Science Foundation CBET-0955172 to JTB, and National Institute of Health HL110328 to JTB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. Lastly, we thank the anonymous reviewers for their helpful suggestions.

## References

1. Thiery JP (2003) Epithelial-mesenchymal transitions in development and pathologies. *Curr Opin Cell Biol* 15: 740-6. 813  
814  
815
2. Stahl PJ, Felsen D (2001) Transforming growth factor-beta, basement membrane, and epithelial-mesenchymal transdifferentiation: implications for fibrosis in kidney disease. *Am J Pathol* 159: 1187-92. 816  
817  
818
3. Eisenberg LM, Markwald RR (1995) Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ Res* 77: 1-6. 819  
820
4. Zavadil J, Böttinger EP (2005) Tgf-beta and epithelial-to-mesenchymal transitions. *Oncogene* 24: 5764-74. 821  
822
5. Xu J, Lamouille S, Derynck R (2009) Tgf-beta-induced epithelial to mesenchymal transition. *Cell Res* 19: 156-72. 823  
824
6. Medici D, Hay ED, Goodenough DA (2006) Cooperation between snail and lef-1 transcription factors is essential for tgf-beta1-induced epithelial-mesenchymal transition. *Mol Biol Cell* 17: 1871-9. 825  
826  
827
7. Medici D, Hay ED, Olsen BR (2008) Snail and slug promote epithelial-mesenchymal transition through beta-catenin-t-cell factor-4-dependent expression of transforming growth factor-beta3. *Mol Biol Cell* 19: 4875-87. 828  
829  
830  
831
8. Kim K, Lu Z, Hay ED (2002) Direct evidence for a role of beta-catenin/lef-1 signaling pathway in induction of emt. *Cell Biol Int* 26: 463-76. 832  
833  
834
9. Jiang YG, Luo Y, He Dl, Li X, Zhang Ll, et al. (2007) Role of wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha. *Int J Urol* 14: 1034-9. 835  
836  
837  
838
10. Huber MA, Azoitei N, Baumann B, Grünert S, Sommer A, et al. (2004) Nf-kappab is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest* 114: 569-81. 839  
840  
841  
842
11. Hardy KM, Booth BW, Hendrix MJC, Salomon DS, Strizzi L (2010) Erbb/egf signaling and emt in mammary development and breast cancer. *J Mammary Gland Biol Neoplasia* 15: 191-9. 843  
844  
845
12. Ahmed S, Nawshad A (2007) Complexity in interpretation of embryonic epithelial-mesenchymal transition in response to transforming growth factor-beta signaling. *Cells Tissues Organs* 185: 131-45. 846  
847  
848
13. Chung SW, Miles FL, Sikes RA, Cooper CR, Farach-Carson MC, et al. (2009) Quantitative modeling and analysis of the transforming growth factor beta signaling pathway. *Biophys J* 96: 1733-50. 849  
850  
851

14. Vilar JMG, Jansen R, Sander C (2006) Signal processing in the tgf-beta superfamily ligand-receptor network. *PLoS Comput Biol* 2: e3. 852  
853
15. Cirit M, Haugh JM (2012) Data-driven modelling of receptor tyrosine kinase signalling networks quantifies receptor-specific potencies of pi3k- and ras-dependent erk activation. *Biochem J* 441: 77-85. 854  
855  
856
16. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* 7: e1001099. 857  
858  
859  
860
17. Bailey JE (2001) Complex biology with no parameters. *Nat Biotechnol* 19: 503-4. 861  
862
18. Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. *Science* 342: 604-7. 863  
864  
865
19. Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). *Biotechnol J* 5: 768-80. 866  
867  
868
20. Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol* 7: e1002263. 869  
870  
871
21. Luan D, Szlam F, Tanaka KA, Barie PS, Varner JD (2010) Ensembles of uncertain mathematical models can identify network response to therapeutic interventions. *Mol Biosyst* 6: 2272-86. 872  
873  
874
22. Kalita MK, Sargsyan K, Tian B, Paulucci-Holthauzen A, Najm HN, et al. (2011) Sources of cell-to-cell variability in canonical nuclear factor-kb (nf-kb) signaling pathway inferred from single cell dynamic images. *J Biol Chem* 286: 37741-57. 875  
876  
877  
878
23. Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, et al. (2011) Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics* 12: 125. 879  
880  
881
24. Creixell P, Schoof EM, Erler JT, Linding R (2012) Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* 30: 842-8. 882  
883
25. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* 20: 370-375. 884  
885  
886  
887
26. Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, et al. (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol Syst Biol* 5: 239. 888  
889  
890

27. Tasseeff R, Nayak S, Song SO, Yen A, Varner JD (2011) Modeling and  
analysis of retinoic acid induced differentiation of uncommitted  
precursor cells. *Integr Biol (Camb)* 3: 578–591.  
891  
892  
893
28. Gadkar KG, Varner J, Doyle FJ 3rd (2005) Model identification of signal  
transduction networks from data using a state regulator problem. *Syst  
Biol (Stevenage)* 2: 17–30.  
894  
895  
896
29. Lian Yg, Zhou Qg, Zhang Yj, Zheng Fl (2011) Vegf ameliorates  
tubulointerstitial fibrosis in unilateral ureteral obstruction mice via  
inhibition of epithelial-mesenchymal transition. *Acta Pharmacol Sin* 32:  
1513–21.  
897  
898  
899  
900
30. Hemavathy K, Ashraf SI, Ip YT (2000) Snail/slug family of repressors:  
slowly going into the fast lane of development and cancer. *Gene* 257:  
1–12.  
901  
902  
903
31. Hemavathy K, Guru SC, Harris J, Chen JD, Ip YT (2000) Human slug  
is a repressor that localizes to sites of active transcription. *Mol Cell Biol*  
20: 5087–95.  
904  
905  
906
32. Dhasarathy A, Phadke D, Mav D, Shah RR, Wade PA (2011) The  
transcription factors snail and slug activate the transforming growth  
factor-beta signaling pathway in breast cancer. *PLoS One* 6: e26514.  
907  
908  
909
33. Guaita S, Puig I, Franci C, Garrido M, Dominguez D, et al. (2002) Snail  
induction of epithelial to mesenchymal transition in tumor cells is  
accompanied by muc1 repression and zeb1 expression. *J Biol Chem* 277:  
39209–16.  
910  
911  
912  
913
34. Jackstadt R, Röh S, Neumann J, Jung P, Hoffmann R, et al. (2013) Ap4  
is a mediator of epithelial-mesenchymal transition and metastasis in  
colorectal cancer. *J Exp Med* 210: 1331–50.  
914  
915  
916
35. Eastman Q, Grosschedl R (1999) Regulation of lef-1/tcf transcription  
factors by wnt and other signals. *Curr Opin Cell Biol* 11: 233–40.  
917  
918
36. Park SY, Lee HE, Li H, Shipitsin M, Gelman R, et al. (2010)  
Heterogeneity for stem cell-related markers according to tumor subtype  
and histologic stage in breast cancer. *Clin Cancer Res* 16: 876–87.  
919  
920  
921
37. Tasseeff R, Nayak S, Salim S, Kaushik P, Rizvi N, et al. (2010) Analysis of  
the molecular networks in androgen dependent and independent prostate  
cancer revealed fragile and robust subsystems. *PLoS One* 5: e8864.  
922  
923  
924
38. Aramburu J, Yaffe MB, López-Rodríguez C, Cantley LC, Hogan PG,  
et al. (1999) Affinity-driven peptide selection of an nfat inhibitor more  
selective than cyclosporin a. *Science* 285: 2129–33.  
925  
926  
927
39. Dhimolea E, Maffini MV, Soto AM, Sonnenschein C (2010) The role of  
collagen reorganization on mammary epithelial morphogenesis in a 3d  
culture model. *Biomaterials* 31: 3622–3630.  
928  
929  
930

40. Butcher JT, Penrod AM, García AJ, Nerem RM (2004) Unique morphology and focal adhesion development of valvular endothelial cells in static and fluid flow environments. *Arterioscler Thromb Vasc Biol* 24: 1429-34. 931  
932  
933  
934
41. Hong JP, Li XM, Li MX, Zheng FL (2013) Vegf suppresses epithelial-mesenchymal transition by inhibiting the expression of smad3 and mir-192 a smad3-dependent microrna. *Int J Mol Med* 31: 1436-42. 935  
936  
937
42. Nagy JA, Dvorak AM, Dvorak HF (2007) Vegf-a and the induction of pathological angiogenesis. *Annu Rev Pathol* 2: 251-75. 938  
939
43. Ferrara N (2002) Vegf and the quest for tumour angiogenesis factors. *Nat Rev Cancer* 2: 795-803. 940  
941
44. Willis BC, Borok Z (2007) Tgf-beta-induced emt: mechanisms and implications for fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol* 293: L525-34. 942  
943  
944
45. Polyak K, Weinberg RA (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* 9: 265–273. 945  
946  
947
46. Strauss R, Li ZY, Liu Y, Beyer I, Persson J, et al. (2011) Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLoS One* 6: e16186. 948  
949  
950
47. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10: 515-27. 951  
952  
953
48. Welch-Reardon KM, Wu N, Hughes CCW (2014) A role for partial endothelial-mesenchymal transitions in angiogenesis? *Arterioscler Thromb Vasc Biol* . 954  
955  
956
49. Zajchowski DA, Bartholdi MF, Gong Y, Webster L, Liu HL, et al. (2001) Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells. *Cancer Res* 61: 5168-78. 957  
958  
959
50. Mancini M, Toker A (2009) Nfat proteins: emerging roles in cancer progression. *Nat Rev Cancer* 9: 810–820. 960  
961
51. Singh G, Singh SK, König A, Reutlinger K, Nye MD, et al. (2010) Sequential activation of nfat and c-myc transcription factors mediates the tgf-beta switch from a suppressor to a promoter of cancer cell proliferation. *J Biol Chem* 285: 27241-50. 962  
963  
964  
965
52. Debnath J, Muthuswamy SK, Brugge JS (2003) Morphogenesis and oncogenesis of mcf-10a mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 30: 256–268. 966  
967  
968

53. Pearson GW, Hunter T (2007) Real-time imaging reveals that  
noninvasive mammary epithelial acini can contain motile cells. *J Cell  
Biol* 179: 1555-67. 969  
970  
971
54. O'Brien LE, Tang K, Kats ES, Schutz-Geschwender A, Lipschutz JH,  
et al. (2004) Erk and mmpps sequentially regulate distinct stages of  
epithelial tubule development. *Dev Cell* 7: 21–32. 972  
973  
974
55. Debnath J, Brugge JS (2005) Modelling glandular epithelial cancers in  
three-dimensional cultures. *Nat Rev Cancer* 5: 675-88. 975  
976
56. Moles CG, Mendes P, Banga JR (2003) Parameter estimation in  
biochemical pathways: a comparison of global optimization methods.  
*Genome Res* 13: 2467-74. 977  
978  
979
57. Rodriguez-Fernandez M, Rehberg M, Kremling A, Banga JR (2013)  
Simultaneous model discrimination and parameter estimation in  
dynamic models of cellular systems. *BMC Syst Biol* 7: 76. 980  
981  
982
58. Villaverde AF, Banga JR (2014) Reverse engineering and identification  
in systems biology: strategies, perspectives and challenges. *J R Soc  
Interface* 11: 20130505. 983  
984  
985
59. Choi M, Shi J, Jung SH, Chen X, Cho KH (2012) Attractor landscape  
analysis reveals feedback loops in the p53 network that control the  
cellular response to dna damage. *Sci Signal* 5: ra83. 986  
987  
988
60. Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, et al.  
(2012) Cellnopt: a flexible toolkit to train protein signaling networks to  
data using multiple logic formalisms. *BMC Syst Biol* 6: 133. 989  
990  
991
61. Hyduke DR, Palsson BØ (2010) Towards genome-scale signalling  
network reconstructions. *Nat Rev Genet* 11: 297-307. 992  
993
62. Wayman J, Varner J (2013) Biological systems modeling of metabolic  
and signaling networks. *Curr Opin Chem Eng* 2: 365 - 372. 994  
995
63. Sainani KL (2012) Meet the skeptics: Why some doubt biomedical  
models - and what it takes to win them over. *Biomedical Computation  
Review* : 12 - 18. 996  
997  
998
64. Lee MW, Vassiliadis VS, Park JM (2009) Individual-based and  
stochastic modeling of cell population dynamics considering substrate  
dependency. *Biotechnol Bioeng* 103: 891-9. 999  
1000  
1001
65. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic  
contributions to stochasticity in gene expression. *Proc Natl Acad Sci U  
S A* 99: 12795-800. 1002  
1003  
1004
66. Handl J, Kell DB, Knowles J (2007) Multiobjective optimization in  
bioinformatics and computational biology. *IEEE/ACM Trans Comput  
Biol Bioinform* 4: 279-92. 1005  
1006  
1007

67. Vega S, Morales AV, Ocaña OH, Valdés F, Fabregat I, et al. (2004) Snail 1008  
blocks the cell cycle and confers resistance to cell death. *Genes Dev* 18: 1009  
1131-43. 1010
68. Nawshad A, Hay ED (2003) Tgfbeta3 signaling activates transcription of 1011  
the lef1 gene to induce epithelial-mesenchymal transformation during 1012  
mouse palate development. *J Cell Biol* 163: 1291-301. 1013
69. Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, et al. 1014  
(2000) The transcription factor snail controls epithelial-mesenchymal 1015  
transitions by repressing e-cadherin expression. *Nat Cell Biol* 2: 76-83. 1016
70. Arce L, Pate KT, Waterman ML (2009) Groucho binds two conserved 1017  
regions of lef-1 for hdac-dependent repression. *BMC Cancer* 9: 159. 1018
71. Grumolato L, Liu G, Haremaiki T, Mungamuri SK, Mong P, et al. (2013) 1019  
?-catenin-independent activation of tcf1/lef1 in human hematopoietic 1020  
tumor cells through interaction with atf2 transcription factors. *PLoS* 1021  
*Genet* 9: e1003603. 1022
72. Zhou BP, Deng J, Xia W, Xu J, Li YM, et al. (2004) Dual regulation of 1023  
snail by gsk-3beta-mediated phosphorylation in control of 1024  
epithelial-mesenchymal transition. *Nat Cell Biol* 6: 931-40. 1025
73. Larue L, Bellacosa A (2005) Epithelial-mesenchymal transition in 1026  
development and cancer: role of phosphatidylinositol 3' kinase/akt 1027  
pathways. *Oncogene* 24: 7443-54. 1028
74. Wu Y, Deng J, Rychahou PG, Qiu S, Evers BM, et al. (2009) 1029  
Stabilization of snail by nf-kappab is required for inflammation-induced 1030  
cell migration and invasion. *Cancer Cell* 15: 416-28. 1031
75. Sullivan NJ, Sasser AK, Axel AE, Vesuna F, Raman V, et al. (2009) 1032  
Interleukin-6 induces an epithelial-mesenchymal transition phenotype in 1033  
human breast cancer cells. *Oncogene* 28: 2940-7. 1034
76. Niessen K, Fu Y, Chang L, Hoodless PA, McFadden D, et al. (2008) Slug 1035  
is a direct notch target required for initiation of cardiac cushion 1036  
cellularization. *J Cell Biol* 182: 315-25. 1037
77. Bullock MD, Sayan AE, Packham GK, Mirnezami AH (2012) Micrornas: 1038  
critical regulators of epithelial to mesenchymal (emt) and mesenchymal 1039  
to epithelial transition (met) in cancer progression. *Biol Cell* 104: 3-12. 1040
78. Linding R, Jensen LJ, Ostheimer GJ, van Vugt MATM, Jørgensen C, 1041  
et al. (2007) Systematic discovery of in vivo phosphorylation networks. 1042  
*Cell* 129: 1415-26. 1043
79. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, et al. (2009) 1044  
String 8—a global view on proteins and their functional interactions in 1045  
630 organisms. *Nucleic Acids Res* 37: D412-6. 1046

80. Derynck R, Zhang YE (2003) Smad-dependent and smad-independent pathways in tgf-beta family signalling. *Nature* 425: 577-84. 1047  
1048
81. Massagué J, Seoane J, Wotton D (2005) Smad transcription factors. 1049  
1050  
*Genes Dev* 19: 2783-810.
82. Nawshad A, Medici D, Liu CC, Hay ED (2007) Tgfbeta3 inhibits 1051  
e-cadherin gene expression in palate medial-edge epithelial cells through 1052  
a smad2-smad4-lef1 transcription complex. *J Cell Sci* 120: 1646-53. 1053
83. Medici D, Potenta S, Kalluri R (2011) Transforming growth factor-?2 1054  
promotes snail-mediated endothelial-mesenchymal transition through 1055  
convergence of smad-dependent and smad-independent signalling. 1056  
*Biochem J* 437: 515–520. 1057
84. Gatza CE, Oh SY, Blobe GC (2010) Roles for the type iii tgf-beta 1058  
receptor in human cancer. *Cell Signal* 22: 1163-74. 1059
85. Hindmarsh A, Brown P, Grant K, Lee S, Serban R, et al. (2005) 1060  
Sundials: Suite of nonlinear and differential/algebraic equation solvers. 1061  
*ACM Transactions on Mathematical Software* 31: 363-396. 1062
86. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy 1063  
estimation and model selection. In: International joint Conference on 1064  
artificial intelligence. Citeseer, volume 14, pp. 1137–1145. 1065
87. Chang H, Park C, Parvin B (2007) Quantitative representation of 1066  
three-dimensional cell culture models. In: Proceedings of the 2007 IEEE 1067  
International Symposium on Biomedical Imaging: From Nano to Macro, 1068  
Washington, DC, USA, April 12-16, 2007. pp. 89–92. 1069  
doi:10.1109/ISBI.2007.356795. URL 1070  
<http://dx.doi.org/10.1109/ISBI.2007.356795>. 1071
88. Polizzotti L, Basak O, Bjornsson C, Shubert K, Yener B, et al. (2012) 1072  
Novel image analysis approach quantifies morphological characteristics of 1073  
3d breast culture acini with varying metastatic potentials. *J Biomed 1074  
Biotech* 2012: 1-16. 1075
89. Nawshad A, Lagamba D, Polad A, Hay ED (2005) Transforming growth 1076  
factor-beta signaling during epithelial-mesenchymal transformation: 1077  
implications for embryogenesis and tumor metastasis. *Cells Tissues 1078  
Organs* 179: 11-23. 1079
90. Miettinen PJ, Ebner R, Lopez AR, Derynck R (1994) Tgf-beta induced 1080  
transdifferentiation of mammary epithelial cells to mesenchymal cells: 1081  
involvement of type i receptors. *J Cell Biol* 127: 2021–2036. 1082
91. Fan JM, Ng YY, Hill PA, Nikolic-Paterson DJ, Mu W, et al. (1999) 1083  
Transforming growth factor-beta regulates tubular 1084  
epithelial-myofibroblast transdifferentiation in vitro. *Kidney Int* 56: 1085  
1455–1467. 1086

92. Hales AM, Schulz MW, Chamberlain CG, McAvoy JW (1994) Tgf-beta 1 induces lens cells to accumulate alpha-smooth muscle actin, a marker for subcapsular cataracts. *Curr Eye Res* 13: 885–890. 1087
93. Kasai H, Allen JT, Mason RM, Kamimura T, Zhang Z (2005) Tgf-beta1 induces human alveolar epithelial to mesenchymal cell transition (emt). *Respir Res* 6: 56. 1088
94. Willis BC, Liebler JM, Luby-Phelps K, Nicholson AG, Crandall ED, et al. (2005) Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *Am J Pathol* 166: 1321–1332. 1089
95. Dennler S, Goumans MJ, ten Dijke P (2002) Transforming growth factor beta signal transduction. *J Leukoc Biol* 71: 731–40. 1090
96. Kavsak P, Rasmussen RK, Causing CG, Bonni S, Zhu H, et al. (2000) Smad7 binds to smurf2 to form an e3 ubiquitin ligase that targets the tgf beta receptor for degradation. *Mol Cell* 6: 1365–1375. 1091
97. Bonni S, Wang HR, Causing CG, Kavsak P, Stroschein SL, et al. (2001) Tgf-beta induces assembly of a smad2-smurf2 ubiquitin ligase complex that targets snon for degradation. *Nat Cell Biol* 3: 587–595. 1092
98. Ju W, Ogawa A, Heyer J, Nierhof D, Yu L, et al. (2006) Deletion of smad2 in mouse liver reveals novel functions in hepatocyte growth and differentiation. *Mol Cell Biol* 26: 654–67. 1093
99. Phanish MK, Wahab NA, Colville-Nash P, Hendry BM, Dockrell MEC (2006) The differential role of smad2 and smad3 in the regulation of pro-fibrotic tgfbeta1 responses in human proximal-tubule epithelial cells. *Biochem J* 393: 601–607. 1094
100. Tian F, DaCosta Byfield S, Parks WT, Yoo S, Felici A, et al. (2003) Reduction in smad2/3 signaling enhances tumorigenesis but suppresses metastasis of breast cancer cell lines. *Cancer Res* 63: 8284–8292. 1095
101. Tian F, Byfield SD, Parks WT, Stuelten CH, Nemani D, et al. (2004) Smad-binding defective mutant of transforming growth factor beta type i receptor enhances tumorigenesis but suppresses metastasis of breast cancer cell lines. *Cancer Res* 64: 4523–4530. 1096
102. Valcourt U, Kowanetz M, Niimi H, Heldin CH, Moustakas A (2005) Tgf-beta and the smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Mol Biol Cell* 16: 1987–2002. 1097
103. Masszi A, Di Ciano C, Sirokmány G, Arthur WT, Rotstein OD, et al. (2003) Central role for rho in tgf-beta1-induced alpha-smooth muscle actin expression during epithelial-mesenchymal transition. *Am J Physiol Renal Physiol* 284: F911–24. 1098

104. Zavadil J, Bitzer M, Liang D, Yang YC, Massimi A, et al. (2001) 1127  
Genetic programs of epithelial cell plasticity directed by transforming 1128  
growth factor-beta. *Proc Natl Acad Sci U S A* 98: 6686–6691. 1129
105. Xie L, Law BK, Chytil AM, Brown KA, Aakre ME, et al. (2004) 1130  
Activation of the erk pathway is required for tgf-beta1-induced emt in 1131  
vitro. *Neoplasia* 6: 603–610. 1132
106. Davies M, Robinson M, Smith E, Huntley S, Prime S, et al. (2005) 1133  
Induction of an epithelial to mesenchymal transition in human immortal 1134  
and malignant keratinocytes by tgf-beta1 involves mapk, smad and ap-1 1135  
signalling pathways. *J Cell Biochem* 95: 918–931. 1136
107. Xie L, Law BK, Aakre ME, Edgerton M, Shyr Y, et al. (2003) 1137  
Transforming growth factor beta-regulated gene expression in a mouse 1138  
mammary gland epithelial cell line. *Breast Cancer Res* 5: R187–R198. 1139
108. Peinado H, Quintanilla M, Cano A (2003) Transforming growth factor 1140  
beta-1 induces snail transcription factor in epithelial cell lines: 1141  
mechanisms for epithelial mesenchymal transitions. *J Biol Chem* 278: 1142  
21113–21123. 1143
109. Peinado H, Portillo F, Cano A (2004) Transcriptional regulation of 1144  
cadherins during development and carcinogenesis. *Int J Dev Biol* 48: 1145  
365–75. 1146
110. Choi J, Park SY, Joo CK (2007) Transforming growth factor-beta1 1147  
represses e-cadherin production via slug expression in lens epithelial cells. 1148  
*Invest Ophthalmol Vis Sci* 48: 2708–2718. 1149
111. Zhang YE (2009) Non-smad pathways in tgf-beta signaling. *Cell Res* 19: 1150  
128–139. 1151
112. Massague J (2003) Integration of smad and mapk pathways: a link and a 1152  
linker revisited. *Genes Dev* 17: 2993–2997. 1153
113. Kretzschmar M, Doody J, Timokhina I, Massagué J (1999) A 1154  
mechanism of repression of tgfbeta/ smad signaling by oncogenic ras. 1155  
*Genes Dev* 13: 804–16. 1156
114. Oft M, Peli J, Rudaz C, Schwarz H, Beug H, et al. (1996) Tgf-beta1 and 1157  
ha-ras collaborate in modulating the phenotypic plasticity and 1158  
invasiveness of epithelial tumor cells. *Genes Dev* 10: 2462–2477. 1159
115. Jungert K, Buck A, von Wichert G, Adler G, König A, et al. (2007) Sp1 1160  
is required for transforming growth factor-beta-induced mesenchymal 1161  
transition and migration in pancreatic cancer cells. *Cancer Res* 67: 1162  
1563–70. 1163

116. Lopez-Casillas F, Riquelme C, Perez-Kato Y, Ponce-Castaneda MV,  
Osses N, et al. (2003) Betaglycan expression is transcriptionally  
up-regulated during skeletal muscle differentiation. cloning of murine  
betaglycan gene promoter and its modulation by myod, retinoic acid,  
and transforming growth factor-beta. *J Biol Chem* 278: 382–390. 1164  
1165  
1166  
1167  
1168
117. Abramoff M, Magelhaes P, Ram S (2004) Image processing with imagej.  
Biophotonics International, 11: 36-42. 1169  
1170
118. Fonseca C, Fleming P, et al. (1993) Genetic algorithms for  
multiobjective optimization: Formulation, discussion and generalization. 1171  
In: Proceedings of the fifth international conference on genetic  
algorithms. Citeseer, volume 423, pp. 416–423. 1172  
1173  
1174
119. Gadkar KG, Doyle FJ 3rd, Crowley TJ, Varner JD (2003) Cybernetic  
model predictive control of a continuous bioreactor with cell recycle.  
Biotechnol Prog 19: 1487-97. 1175  
1176  
1177

**Figure 1.** Model connectivity recreates the core architecture during EMT. The EMT network contains 995 nodes (proteins, mRNA, and genes) interconnected by 1700 interactions. Central to EMT induction, activation of the MAPK cascade occurs through TGF- $\beta$ 1/2 binding which activates the AP-1/Sp1 transcriptional axis. AP-1/Sp1 drives an autocrine response of TGF- $\beta$ 3, which activates the Smad cascade, leading to phenotypic change. Conversely, VEGF-A binding promotes an epithelial phenotype through NFAT activation. Other important signaling pathways not shown but also incorporated include the BMP, Wnt, and PI3K pathways. The complete list of molecular interactions that comprise the model is given in the supplement.

**Figure 2.** Training and validation simulations. The population of EMT models qualitatively captured TGF $\beta$ -induced EMT signaling. (A-I) The population was generated using POETs and trained using 11 different objective functions (41 data sets) taken from Medici *et al.* [7]. The model captured the simulated experiments for 78% of the cases. (J-L) The model populations were also compared against untrained temporal data to measure the effectiveness as a pure prediction. The high predictability can be contributed to the leave-one-out cross validation scheme, objective functions with overlapping data, and multi-objective optimization algorithm.

**Figure 3.** Simulated TGF $\beta$ 1/2 exposure promoted phenotype heterogeneity. Robustness coefficients were used to quantify the effect of perturbations after steady state across the population of data sets. Coefficients with values  $> 1$  ( $< 1$ ) indicated a marker increased (decreased) compared to a base state, while a value of 1 indicated approximately no change following a perturbation. E-cadherin and Vimentin robustness coefficients were used as phenotypic markers. (A) TGF $\beta$ 1/2 perturbation, we isolated 4 distinct parameter sets known to have phenotypically different behaviors. (B) Each region has been represented by a small signaling network. From our analysis, we determined that the differences were a function of downstream transcription factors (phosphorylated-Sp1, and NFAT) within the TGF $\beta$ 2 and VEGFA pathway, respectively. (C) We hypothesized that elevated phosphorylated-Sp1 and NFAT levels could drive phenotype heterogeneity through simultaneous TGF $\beta$ 2 and VEGFA treatment, similar to region three (R3).

**Figure 4.** Simulated response to TGF $\beta$ 1/2 and VEGF-A exposure with and without axis specific inhibitors. Robustness coefficients were used to quantify the shift in population at 48 hrs. (A-C) VEGF-A (50 a.u.) treatment resulted in a population with enhanced epithelial (Q4) properties. This was contrary to the addition of TGF $\beta$ 2 (10 a.u.), which shifted the population towards a mesenchymal phenotype (Q1). Interestingly, the combined effects of TGF $\beta$ 2 and VEGFA was found to increase both ecadherin and vimentin levels, creating a heterogeneous population (Q2). (D-F) To isolate the effect of NFAT, we inhibited NFAT de-phosphorylation in combination with VEGFA. This negated the increase in ecadherin expression and shifted the population towards a mesenchymal phenotype (Q1,Q3). Likewise, combining NFAT inhibition with TGF $\beta$  mitigated all ecadherin expression (Q2) confirming their importance for population heterogeneity. Lastly, combination of TGF $\beta$ 2, VEGFA, and NFAT inhibition nearly mitigated all effects of VEGFA, shifting the heterogeneous population (Q2) towards a mesenchymal phenotype (Q1). In whole, high levels of phosphorylated-Sp1 correlated with vimentin expression, while NFAT was responsible for maintaining E-cadherin expression, although neither were mutually exclusive.

**Figure 5.** Simultaneous TGF- $\beta$ 1/2 and VEGF-A treatment induced phenotype heterogeneity and is dependent upon NFAT activity *in-vitro*. (A) In MCF10A, treatment with (10ng/ml) TGF $\beta$ 2 increased Slug and vimentin, while ecadherin expression was inhibited at both the gene and protein level at 48 hrs. Conversely, VEGFA alone increased both NFATc1 and ecadherin gene expression. Simultaneous TGF $\beta$ 2 (10ng/ml) and VEGFA (50ng/ml) treatment increased Slug, NFATc1, and vimentin expression, while also increasing ecadherin levels via qPCR. (B-C) Immunofluorescence confirmed these results and nuclear co-localization of both phospho-Sp1 and NFAT were found dependent upon TGF $\beta$ 2 and VEGFA, respectively. (D) To isolate the effect of NFAT, treatment of VEGFA (50ng/ml) and VIVIT (10 $\mu$ M) reduced ecadherin expression at 48hrs (control-dashed line). Similarly, combined TGF $\beta$ 2, VEGFA and VIVIT treatment increased Slug and vimentin expression, while inhibiting ecadherin levels via qPCR. (E) These findings were confirmed via immunofluorescence as the VIVIT peptide inhibited ecadherin and nuclear localization of NFATc1 in all three cases. (F) Quantitative flow cytometry also confirmed this trend. Similar experiments in DLD1 followed a similar trend (supplement). Magnification, 40x. Scale bars: 50 $\mu$ m. C=Control, T=TGF $\beta$ 2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ( $p < 0.05$ ).

**Figure 6.** Ductal branching is dependent upon phenotype heterogeneity within MCF10A in 3-D culture. MCF10A and DLD1 were formed into spheroids overnight and explanted to a collagen gel for 72 hrs. For compaction and tubular assays, cells were embedded into collagen gels for 72 hrs, and the extent of tubulogenesis was measured at 7 days. (A-D) Within MCF10A, TGF $\beta$ 2 (10ng/ml) enhanced invasion and contractile properties while, VEGFA (50ng/ml) promoted increased migration. TGF $\beta$ 2 with VEGFA significantly increased migration, while limiting with compaction. VIVIT (10 $\mu$ M) in combination with VEGFA and TGF $\beta$ 2 decreased migration and compaction, while increasing invasion. (D) Likewise, cell morphology (circularity index) correlated with both invasion and compaction in MCF10A. (E-F) The size of tubular structures (acini) also increased significantly upon addition of VEGFA, while the number of ductal branches was most significant upon simultaneous TGF $\beta$ 2 and VEGFA treatment (Red-Ecadherin, Green-Factin, Blue-Nuclear). DLD1 cells followed a similar trend, although the degree of migration, invasion, and compaction was less significant. In addition, no tubular structures were identified during the 7 day tubulogenesis endpoints. Scale bars: 500 $\mu$ m, 1000 $\mu$ m, 250 $\mu$ m, and 80 $\mu$ m, respectively. C=Control, T=TGF $\beta$ 2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ( $p < 0.05$ ). Boxes in the left-most panel identify regions identified by arrows that were then imaged in greater zoom in the panel immediately below. The box diagram was not repeated for arrows in the other panels for clarity, but the same method was applied.

## Supplemental Materials and Methods

1178

### EMT model network architecture.

1179

TGF $\beta$  is a major inducer of EMT in development, fibrosis, and carcinogenesis with different isoforms mediating various effects depending on specific cellular context [89]. TGF $\beta$ 1 was first described as an inducer of EMT in normal mammary epithelial cells [90] and has since been shown to mediate EMT in vitro in a number of different epithelial cells, including renal proximal tubular, DLD1 colon carcinoma, and most recently alveolar epithelial cells [91–94]. TGF $\beta$  signaling occurs through the Smad pathway in which signals are transduced by transmembrane serine/threonine kinase type I (ALK5) and type II (TGF $\beta$ RII) receptors. To increase ligand affinity, betaglycan (TGF $\beta$ RIII) can also interact with TGF $\beta$ RI,II [84]. Upon TGF $\beta$  stimulation, the receptors are internalized into early endosomes where Smad anchor for receptor activation (SARA) modulates formation of complexes with (R-Smad) Smad2 or Smad3. Smad2 and Smad3 are then phosphorylated at serine residues by the type I receptor [81]. Phosphorylation induces their association with (Co-Smad) Smad4 and translocation to the nucleus where they interact with other transcription factors to regulate the transcription of TGF $\beta$  responsive genes, including alpha-smooth muscle actin, collagen1A2, vimentin, fibronectin, and plasminogen activator inhibitor-1 (PAI-1) by interacting with Smad-binding elements [80,95]. To regulate TGF $\beta$  signaling, Smurf2 (a ubiquitin E3 ligase) can become activated to mediate proteasome dependent degradation of Smad2 or bind with Smad 7 to target TGF $\beta$  receptor for degradation [96,97].

1200

### Cell Type Dependency

1201

Interestingly, differential roles for Smad2 and Smad3 in TGF $\beta$  induced EMT have been demonstrated. For example, using primary cells from mice with hepatocyte-specific double knockout of Smad2 and Smad3, it was demonstrated that Smad3 but not Smad2 was required for a key morphological changes and induction of EMT [98]. In contrast, using human proximal tubular epithelial cells, increased colony growth factor and decreased E-cadherin were Smad3 dependent, increased MMP-2 was Smad2 dependent, while alpha-smooth muscle actin was dependent on both [99]. Furthermore, Smad2 signaling has been demonstrated in cancerous lung epithelial A549 cells [93]. Inhibition of Smad3 signaling decreases the metastatic potential of xenografted breast cancer cell lines [100,101]. Together, these results suggest that the precise Smad pathway activated may depend on the particular cellular context. Regardless, a recent transcriptomic analysis of TGF $\beta$  induced EMT in normal mouse and human epithelial cells using a dominant negative approach demonstrated that Smad signaling was critical for regulation of all tested target genes [102].

1202

1203

1204

1205

1206

1207

1208

1209

1210

1211

1212

1213

1214

1215

1216

### Smad-Independent EMT Induction Pathways

1217

Although less well established than the Smad-dependent pathways in the induction of EMT, there is substantial evidence for TGF $\beta$  activation of

1218

1219

Smad-independent signaling in some aspects of this process. These include RhoA, MAPK, PI3 kinase, and Notch signaling pathway, which have mostly been studied in-vitro. For example, the small GTPase RhoA is involved in TGF $\beta$  induced EMT in a number of cell types including NMuMG mammary epithelial cells and mink lung epithelial (Mv1Lu) cells particularly in the regulation of cytoskeletal and adherens junction rearrangement. In addition to its role in cytoskeletal remodeling, Rho has been shown to activate the alpha-smooth muscle promoter during EMT in kidney proximal tubular (LLCPK1) cells [103]. TGF $\beta$  has been shown to elevate ERK activity in culture models of EMT (human keratinocytes, NMuMG mammary epithelial cells, and mouse cortical tubule epithelial cells) [104–106]. This ERK activity was required for disassembly of adherens junctions and induction of cell motility [105]. Inhibition of ERK blocked these key morphological changes of EMT in mammary gland epithelial cells [107]. Recent studies in MDCKII and DLD1 colon carcinoma cells revealed that TGF $\beta$ 3 can signal through PI3 kinase. Upon activation, PI3 kinase activates molecules such as ILK and AKT, which can phosphorylate and inactivate GSK3 $\beta$ , a protein that targets both Snail and  $\beta$ -catenin for degradation through the ubiquitin proteasome pathway [7]. Lastly, there is increasing evidence for a role of Notch pathways in regulating EMT, particularly during development [89]. In cardiac cushion EMT, it was found that that Slug is directly up-regulated by Notch in endothelial cells repression of the vascular endothelial cadherin promoter and for promoting migration of transformed endothelial cells. In whole, the distinction between Smad-dependent and Smad-independent mechanisms remains difficult to interpret due to the significant cross talk between these pathways. In most cases, stimulation of these cooperative pathways provides the context for induction and specification of EMT within a particular tissue/cell type, with Smads representing the dominant pathway, which in some instances may be necessary but not sufficient for induction of full EMT [4].

### Master Transcription Regulators

Master transcription factors of EMT are thought to be regulated by Snail1 (SNAI1) and Snail2 (SNAI2) (previously known as Snail and Slug, respectively). These transcription factors are zinc finger proteins that function as repressors of E-cadherin transcription in cultured epithelial cells [69], repression of E-cadherin leads to dissolution of adherens junctions. Snail and Slug can be activated by TGF $\beta$  via both Smad-dependent and -independent pathways in a cell type dependent fashion in cultured cells [108]. Differential expression of Snail and Slug is observed in TGF $\beta$  induced EMT in keratinocytes, renal proximal tubular, and mammary epithelial cells, suggesting that they are regulated in a cell specific. Interestingly, key regulatory units of Snail and Slug appear to be regulated by upstream transcription factors AP1 and SP1, respectively. Peinado et al. clearly established a necessary role of Snail in mediating TGF $\beta$ 1 induced EMT in MDCK cells. They demonstrated that the Ras-Raf-MEK-ERK-AP1 signaling pathway could up-regulate synthesis of the E-cadherin repressor molecule Snail (directly binds to Snail promoter) [109]. Likewise, TGF $\beta$ 1

induced the expression of Slug in both lens and other epithelial cells in vitro. The Sp1 binding site in the Slug promoter is largely responsible for TGF $\beta$ 1 induced Slug expression and upstream of MAPK signaling. In addition, the TGF $\beta$ 1 mediated repression of E-cadherin was significantly inhibited by Slug siRNA [110].

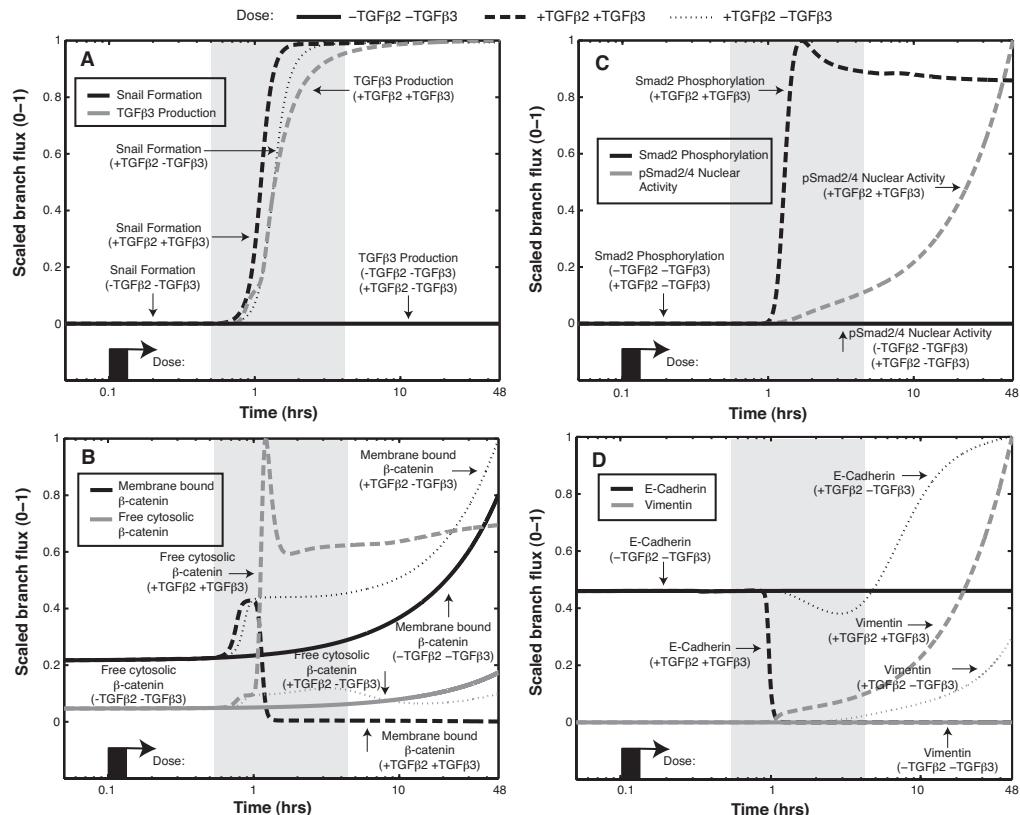
### ERK/MAPK Crosstalk

The ERK-MAPK pathway has been shown to modify TGF $\beta$  signaling at multiple levels. One way is that the activated Ras pathway inhibits the TGF $\beta$  induced nuclear accumulation of Smad2/3, as seen in epithelial cells [111]. ERK kinases have been shown to phosphorylate Smad2 and Smad3 at specific sites in the region linking the MH1 and MH2 domains. These sites are separate from the TGF $\beta$  receptor phosphorylation sites which activate Smad nuclear translocation [112]. The effect of interaction between ERKs and Smads is the subject of some controversy, with data suggesting that such an interaction either enhances or inhibits downstream events. A hyperactive Ras pathway has been shown to effectively counteract the antiproliferative activity of TGF $\beta$  through attenuation of Smad accumulation in the nucleus [113]. In contrast, Ras signals strongly cooperated with Smads for invasion of human carcinoma cells [114]. A second way is through transcriptional regulation. Recently, it was found that Sp1, via transcriptional induction of Vimentin, cooperates with activated Smad complexes in mesenchymal transition and migration of pancreatic cancer cells upon TGF $\beta$  stimulation [115]. Likewise, binding sites at the SP1/Smad3 complex was found to regulate the betaglycan receptors [116]. Taken together, multiple levels of cross-talk (both positive and negative) exist within MAPK and TGF $\beta$ , and may also be cell type dependent.

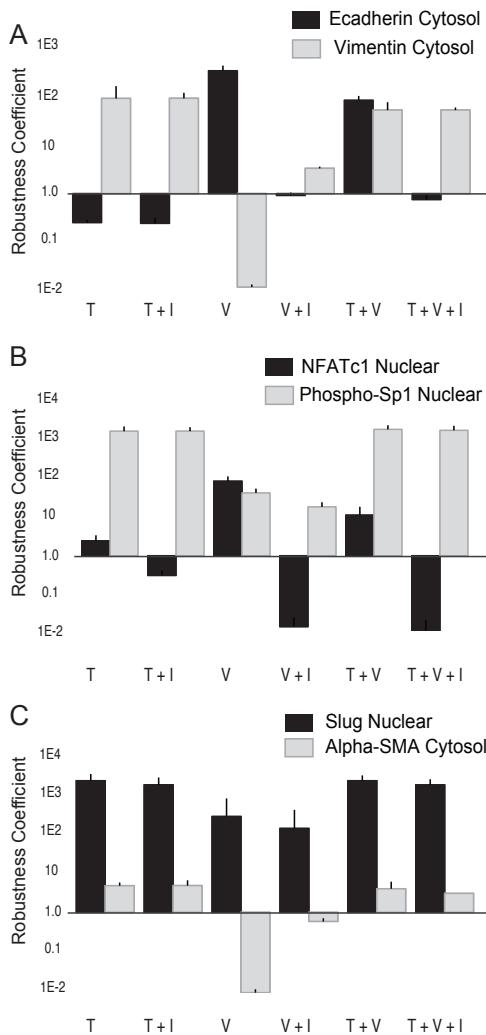
### Analysis of the signal flow through the EMT architecture.

To investigate temporal shifts for key species dominating the EMT response, we calculated the scaled flux through the signaling architecture (Fig. S1). Three modes of operation were simulated to identify distinct behavioral differences: (a) no TGF- $\beta$ 2 (-TGF- $\beta$ 1/2), (b) TGF- $\beta$ 1/2 stimulation (+TGF- $\beta$ 2), and (c) TGF- $\beta$ 1/2 stimulation while blocking the autocrine response of TGF- $\beta$ 3 (+TGF- $\beta$ 1/2 + TGF- $\beta$ 3Ab). Following TGF- $\beta$ 1/2 stimulation, AP1/SP1 rapidly stimulates Snail/Slug expression (Fig. S1A) within 1 hr. During this phase, Snail/Slug repress E-cadherin expression, which in turn reduces the interaction between E-cadherin and  $\beta$ -catenin, and the accumulation of free cytosolic  $\beta$ -catenin ensues. Conversely, in the presence of a TGF- $\beta$ 3 neutralizing antibody, membrane bound  $\beta$ -catenin increased after 10 hrs, while low levels of free cytosolic  $\beta$ -catenin accumulated between 1-5 hrs (Fig. S2A,B). Accumulation of free  $\beta$ -catenin is critical for complexing with TCF4 to rapidly produce the autocrine response of TGF- $\beta$ 3 within 1-10 hrs. Following TGF- $\beta$ 3 expression, formation and spatial relocation of the Smad complexes occurs in a time dependent manner. Smad2 is phosphorylated within 1 hr and nuclear localization of the pSmad2/4 complex dramatically increases after 10 hours (Fig. S3).

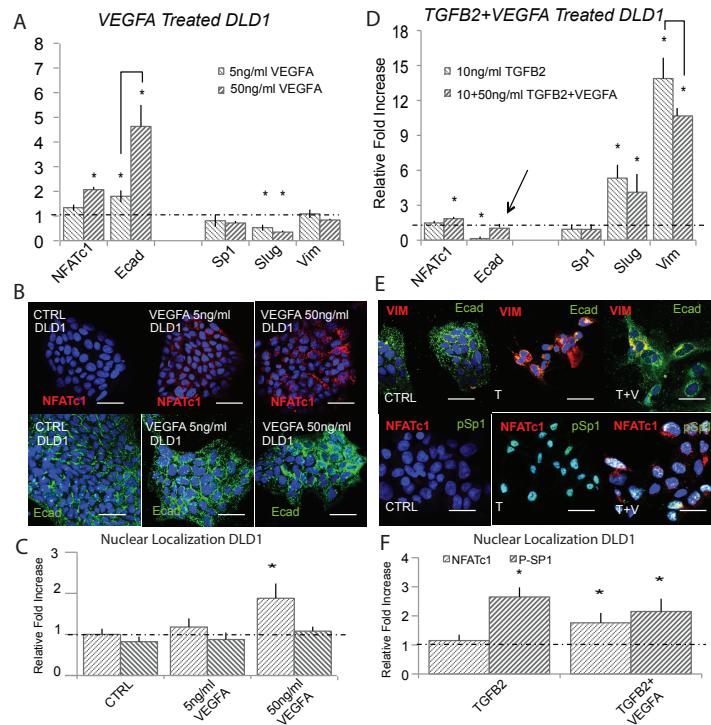
S1C). This timely formation is critical for completion of the EMT process. At 1308  
steady state, the E-cadherin complex was maintained at a basal epithelial level 1309  
with no expression of Vimentin (Fig. S1D). During TGF- $\beta$ 1/2 stimulation, both 1310  
MAPK and Smad act synergistically to repress the E-cadherin complex within 1 1311  
hr. This is followed by increase of Vimentin at ~ 1hr, while exponentially 1312  
increasing around 10 hrs. Both species elevate within 1 hr, however the complex 1313  
does not significantly form until ~ 10 hrs. When blocking TGF- $\beta$ 3, Snail/Slug 1314  
downregulates the E-cadherin complex between 5-10 hrs, but cannot complete 1315  
the transformation due to the loss of Smad2 phosphorylation (Fig. S1D). 1316



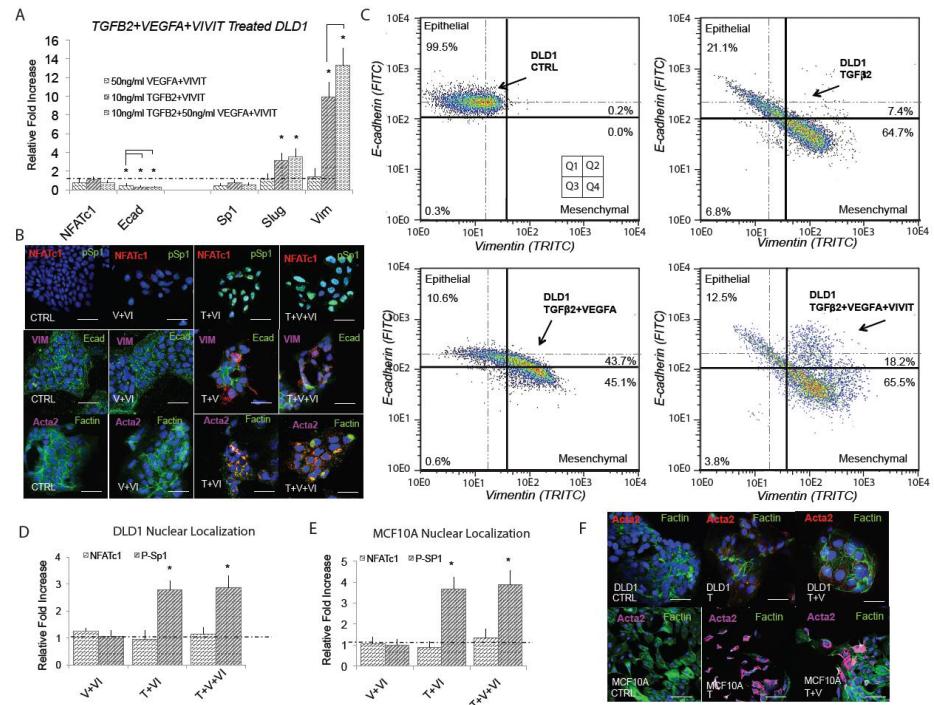
**Figure S1.** Signal flow analysis of key species at steady state, TGF- $\beta$ 1/2 stimulation, and blocking the TGF- $\beta$ 3 autocrine response. (A) The MAPK cascade is directly responsible for rapid expression of Snail and downstream TGF- $\beta$ 3 formation (1 hrs). (B) TGF- $\beta$ 1/2 reduces  $\beta$ -catenin, allowing rapid free-cytosolic  $\beta$ -catenin to accumulate (1hr). Blocking TGF $\beta$ 3 increases membrane bound  $\beta$ -catenin (10hr). (C) TGF- $\beta$ 3 activates the Smad cascade. Nuclear localization of the pSmad2/4 complex (10 hrs) is dependent upon both the phosphorylation of Smad2 (1 hrs) and complexing with Smad4 (5 hrs). (D) TGF- $\beta$ 1/2 rapidly reduces the E-cadherin complex, while upregulating Vimentin (5-10 hrs). Blocking TGF $\beta$ 3 increases E-cadherin (10 hrs) and Vimentin is significantly reduced.



**Figure S2.** Robustness analysis for key molecular species at  $t = 48$  hrs for combinations of TGF- $\beta 1/2$ , VEGF-A and NFATc1 inhibitors. Robustness coefficients for the indicated species were calculated for  $N \sim 1100$  ensemble members for 48 hrs following the addition of TGF- $\beta 1/2$  (T), TGF- $\beta 1/2$  + NFATc1 inhibitor (T + I), VEGF-A (V), VEGF-A + NFATc1 inhibitor (V + I) and TGF- $\beta 1/2$  + VEGF-A (T+V) + NFATc1 inhibitor (T + V + I). (A) Robustness coefficients for E-cadherin and Vimentin as a function of condition. (B) Robustness coefficients for nuclear localized phosphorylated Sp1 and NFATc1 as a function of condition. (C) Robustness coefficients for nuclear localized Slug and  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA) as a function of condition. In each case the error bars denote one-standard deviation of robustness coefficient calculated over the model ensemble. C=Control, T=TGF $\beta 2$ , V=VEGFA, VI= NFAT inhibitor (VIVIT).



**Figure S3.** VEGF-A attenuates TGF- $\beta$ 1/2 to induce phenotype heterogeneity in DLD1. (A) In DLD1, we found that 5ng/ml of VEGFA increased NFATc1 and E-cadherin gene expression via qPCR and 50ng/ml potentiated this effect at 48 hrs. (B - C) These findings were confirmed at the protein level via immunofluorescence, as ecadherin levels and nuclear localization of NFATc1 increased. (D) Treatment with (10ng/ml) TGF $\beta$ 2 resulted in mesenchymal transformation as measured via qPCR against target genes Slug, ecadherin, vimentin, Sp1, and NFATc1. (E - F) Immunofluorescence and nuclear localization revealed a strong presence of phospho-Sp1. (G) Combination of VEGFA (50ng/ml) and TGF $\beta$ 2 (10ng/ml) treatment resulted in increased Slug, NFATc1, and vimentin expression, while also increasing ecadherin levels compared to control. (H) Immunofluorescence confirmed these results, as both ecadherin and vimentin levels were elevated. (I) A significant increase in nuclear localization of both NFATc1 and phospho-Sp1 were also found. Magnification, 40x. Scale bars: 50 $\mu$ m. C=Control, T=TGF $\beta$ 2 , V=VEGFA, VI=NFMAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ( $p < 0.05$ ).



**Figure S4.** E-cadherin expression is dependent upon NFAT activity in DLD1. (A) Treatment with VEGFA (50ng/ml) and NFAT inhibitory peptide VIVIT ( $10\mu\text{M}$ ) resulted in significantly reduced ecadherin expression (qRT-PCR at 48hrs). Addition of TGF $\beta$ 2 (10ng/ml) and VIVIT resulted in increased Slug and vimentin expression, while inhibiting ecadherin levels. Combined TGF $\beta$ 2, VEGFA, and VIVIT treatment resulted in target genes Slug and vimentin expression increased, while inhibiting ecadherin levels. No change in Sp1 or NFATc1 expression was found. (B) These findings were confirmed via immunofluorescence as the VIVIT inhibitors was shown to inhibit ecadherin levels in all three cases. We also found no change in gene or nuclear localization of NFATc1 in all three cases, while phospho-Sp1 was found to increase in both TGF $\beta$  conditions. (C) Quantitative flow cytometry also confirmed this trend. (D,E) TGF $\beta$ 2, VEGFA and VIVIT treatment in DLD1 and MCF10A resulted in no change of Sp1 expression or NFATc1 expression. (F) Likewise, no change in nuclear localization of NFAT in all three cases, however phospho-Sp1 was found to increase in both TGF $\beta$  conditions. Magnification, 40x. Scale bars:  $50\mu\text{m}$ . C=Control, T=TGF $\beta$ 2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ( $p < 0.05$ ).

## Estimation and cross-validation of EMT model parameters.

We used the Pareto Optimal Ensemble Technique (POETs) multiobjective optimization framework in combination with leave-one-out cross-validation to estimate an ensemble of TGF $\beta$ /EMT models. Cross-validation was used to calculate both training and prediction error during the parameter estimation procedure [86]. The 41 intracellular protein and mRNA data-sets used for identification were organized into 11 objective functions. These 11 objective functions were then partitioned, where each partition contained ten training objectives and one validation objective. POETs integrates standard search strategies e.g., Simulated Annealing (SA) or Pattern Search (PS) with a Pareto-rank fitness assignment [19]. Denote a candidate parameter set at iteration  $i + 1$  as  $\mathbf{k}_{i+1}$ . The squared error for  $\mathbf{k}_{i+1}$  for training set  $j$  was defined as:

$$E_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j} \left( \hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k}) \right)^2 \quad (\text{S1})$$

The symbol  $\hat{\mathcal{M}}_{ij}$  denotes scaled experimental observations (from training set  $j$ ) while  $\hat{y}_{ij}$  denotes the scaled simulation output (from training set  $j$ ). The quantity  $i$  denotes the sampled time-index and  $\mathcal{T}_j$  denotes the number of time points for experiment  $j$ . In this study, the experimental data used for model training was typically the band intensity from Western or Northern blots. Band intensity was estimated using the ImageJ software package [117]. The scaled measurement for species  $x$  at time  $i = \{t_1, t_2, \dots, t_n\}$  in condition  $j$  is given by:

$$\hat{\mathcal{M}}_{ij} = \frac{\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}}{\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}} \quad (\text{S2})$$

Under this scaling, the lowest intensity band equaled zero while the highest intensity band equaled one. A similar scaling was defined for the simulation output. By doing this scaling, we trained the model on the relative change in blot intensity, over conditions or time (depending upon the experiment). Thus, when using multiple data sets (possibly from different sources) that were qualitatively similar but quantitatively different e.g., slightly different blot intensities over time or condition, we captured the underlying trends in the scaled data. Additionally, we applied a universal convention of identifying the no expression case as protein value below  $<10^{-3}$ . This is similar to previously published models from our lab. In our studies, Figure 2 identifies experimental data extracted from published Western blots and our simulation results. It is clear from these that 1) the training data included a variety of Western blot data treatments that were effectively zero, and 2) our simulations matched the training data virtually perfectly over time and across multiple biological species. These results validate the power of our simulation scheme, which necessarily includes numerical interpretations of zero. In our Supplemental Figure S9, we present the raw concentrations of our simulations (previously validated from the training data) over hundreds of parameter ensembles, with no “zero” threshold applied. While these values do not have a threshold, we only interpret values below  $10^{-3}$  equivalent to zero. Likewise, our robustness coefficients (Figure 3)

O#	Species (cytosolic protein)	Cell Type	Training	Prediction	Random	Source
O1	LEF1	DLD1 CC, MDCKII, A375 MC	0.73 ± 0.22	0.66 ± 0.18	0.87 ± 0.02	Medici et al., 2008
O2	Vimenin	DLD1 CC, MDCKII, A375 MC	0.96 ± 0.11	1.00 ± 0.15	0.23 ± 0.04	"
O3	TGF beta3	DLD1 CC, MDCKII, A375 MC	1.00 ± 0.13	0.77 ± 0.22	0.61 ± 0.04	"
O4	E-Caderin	DLD1 CC, MDCKII, A375 MC	0.54 ± 0.16	0.00 ± 0.00	0.15 ± 0.03	"
O5	beta-catenin	DLD1 CC, MDCKII, A375 MC	0.99 ± 0.25	0.00 ± 0.00	0.00 ± 0.00	"
O6	TGFBeta3	DLD1 CC, MDCKII, A375 MC	0.89 ± 0.14	0.71 ± 0.12	0.58 ± 0.04	"
O7	Snail	DLD1 CC, MDCKII, A375 MC	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	"
O8	LEF1	DLD1 CC, MDCKII, A375 MC	1.00 ± 0.14	1.00 ± 0.25	0.61 ± 0.05	"
O9	E-Caderin	DLD1 CC, MDCKII, A375 MC	0.96 ± 0.20	0.86 ± 0.19	0.00 ± 0.00	"
O10	Slug	DLD1 CC, MDCKII, A375 MC	1.00 ± 0.20	1.00 ± 0.12	0.00 ± 0.00	"
O11	LEF1	DLD1 CC, MDCKII, A375 MC	1.00 ± 0.15	1.00 ± 0.25	0.86 ± 0.03	"
P#	Species		Simulated	Experimental	Random	Source
P1	E-caderin (mRNA)	MDCKII	0.10 ± 0.01	0.13 ± 0.02	0.27 ± 0.03	Medici et al., 2006
P2	pSmad2	MDCKII	0.58 ± 0.12	0.80 ± 0.04	0.13 ± 0.03	"
P3	LEF1 (mRNA)	MDCKII	0.76 ± 0.05	0.65 ± 0.04	0.18 ± 0.03	"

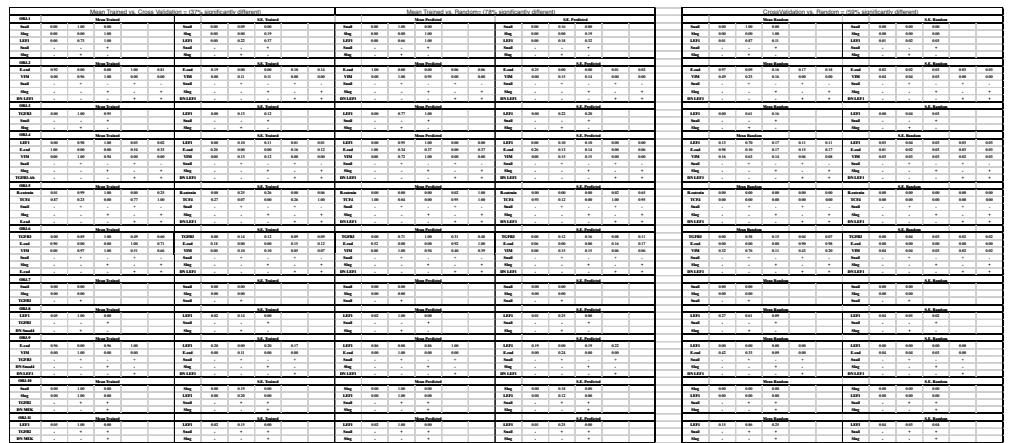
**Figure S5.** Training and prediction values for the 11 TGF- $\beta$  objective functions versus a random parameter control.

(which represent the ratio of integrated areas of the treatment effect over the baseline effect) identify no difference in model output for vimentin (or e-cadherin for that matter) less than  $10^{-3}$ , confirming this interpretation. 1357  
1358  
1359

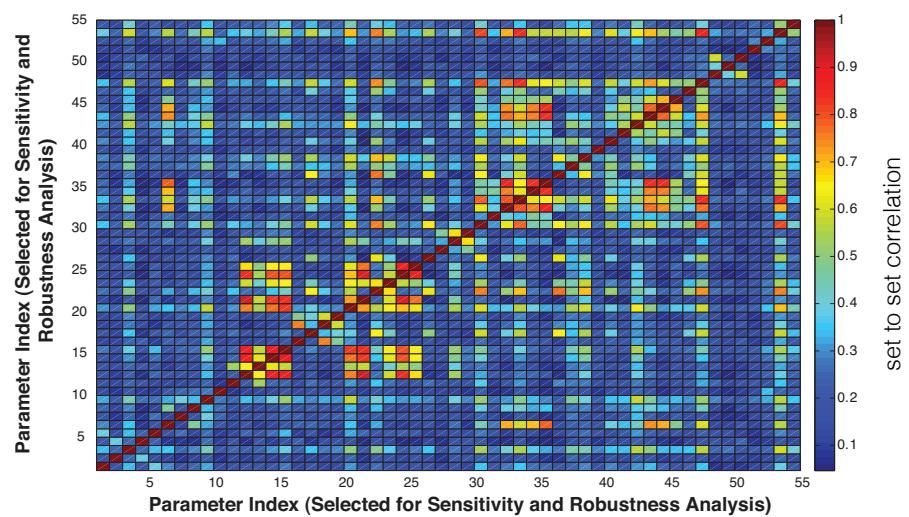
We computed the Pareto rank of  $\mathbf{k}_{i+1}$  by comparing the simulation error at iteration  $i + 1$  against an archive of accepted parameter sets  $\mathbf{K}_i$ . We used the Fonseca and Fleming ranking scheme [118] to estimate the number of parameter sets in the archive that dominate  $\mathbf{k}_{i+1}$ . Parameter sets with increasing rank were progressively further away from the optimal trade-off surface. The parameter set  $\mathbf{k}_{i+1}$  was accepted or rejected by POETs with probability  $\mathcal{P}(\mathbf{k}_{i+1})$ : 1360  
1361  
1362  
1363  
1364  
1365

$$\mathcal{P}(\mathbf{k}_{i+1}) \equiv \exp \{ -\text{rank}(\mathbf{k}_{i+1} | \mathbf{K}_i) / T \} \quad (\text{S3})$$

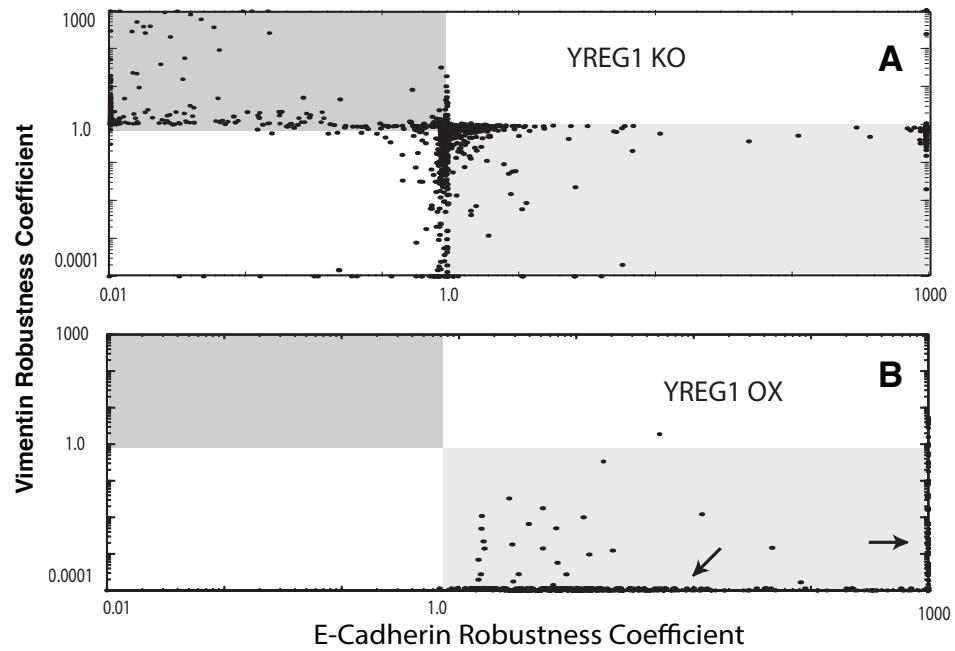
where  $T$  is the annealing temperature and  $\text{rank}(\mathbf{k}_{i+1} | \mathbf{K}_i)$  denotes the Pareto rank for  $\mathbf{k}_{i+1}$ . The annealing temperature was discretized into 10 quanta between  $T_o$  and  $T_f$  and adjusted according to the schedule  $T_k = \beta^k T_0$  where  $\beta$  was defined as  $(T_f/T_o)^{1/10}$ . The initial temperature was  $T_o = n/\log(2)$ , where  $n = 4$  in this study and the final temperature was  $T_f = 0.1$ . The epoch-counter  $k$  was incremented after the addition of 100 members to the ensemble. Thus, as the ensemble grew, the likelihood of accepting parameter sets with a large Pareto rank decreased. To generate parameter diversity, we randomly perturbed each parameter by  $\leq \pm 25\%$  at iteration of the search. In addition, we performed a local pattern search every  $q$ -iterations to minimize the residual for a single random or the worst performing objective function. The local pattern-search algorithm has been described previously [119]. From the 15,000 probable EMT models, we selected  $N = 1093$  models with Pareto rank  $\leq 1$  for subsequent analysis. A quick estimate of the set to set correlation showed that we could expect on order 25% correlation between parameter sets in the ensemble. 1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380



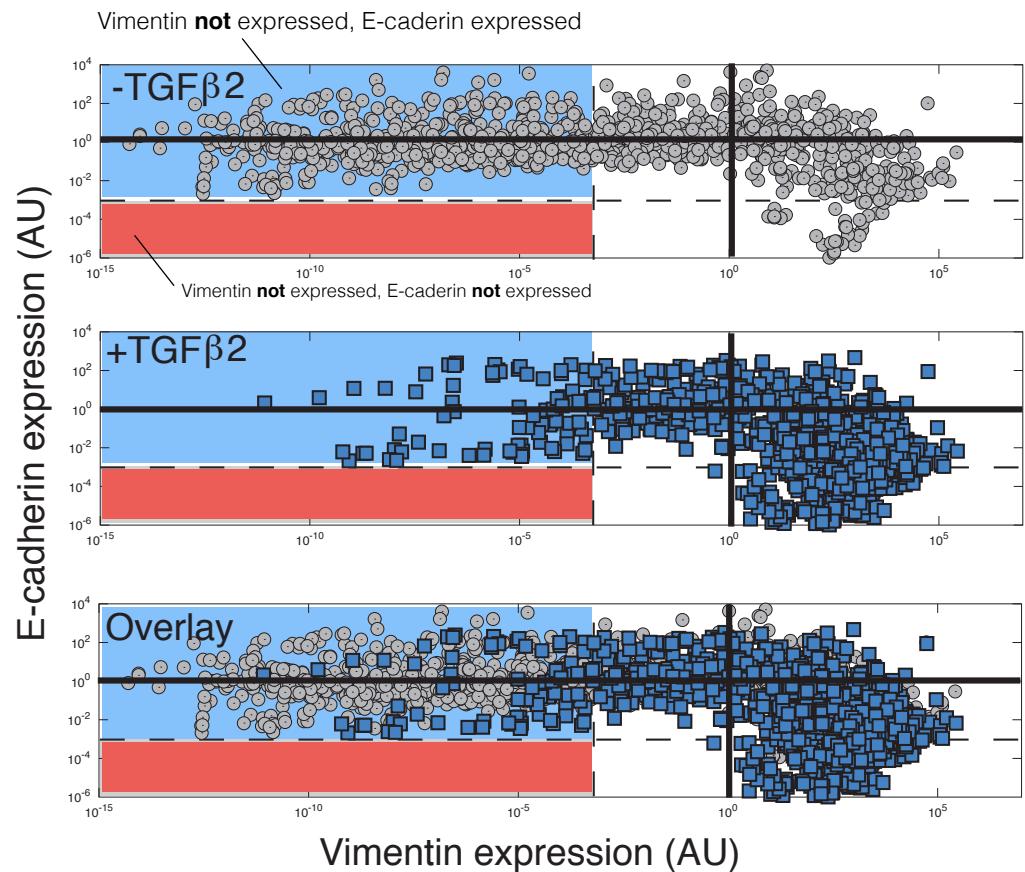
**Figure S6.** Training and prediction values as a function of condition for the 11 TGF- $\beta$  objective functions versus a random parameter control.



**Figure S7.** Parameter set to set correlation for 55 random parameter sets selected from the ensemble. Of the 55 sets selected, the average correlation between sets was less than 25% for greater than 80% of the parameter sets.



**Figure S8.** Robustness of E-cadherin and Vimentin expression to a knockout (A) and overexpression (B) of the hypothetical regulator 1 (YREG1) protein. Robustness coefficients were calculated for each member of the ensemble. Each point represents the response of a single model in the ensemble to either a knockout or overexpression of YREG1.



**Figure S9.** Steady state protein abundance for E-cadherin and Vimentin (AU) as a function of TGF- $\beta$ 1/2 exposure. Top: Overlay of the model population for Vimentin (AU) and E-cadherin (AU) expression in the presence (blue) and absence (gray) of TGF- $\beta$ 1/2. Middle: Vimentin (AU) and E-cadherin (AU) expression in the absence of TGF- $\beta$ 1/2 showed exhibited population heterogeneity. Bottom: Vimentin (AU) and E-cadherin (AU) expression in the presence of TGF- $\beta$ 1/2 moved the centroid of the population toward Vimentin (AU) and away from E-cadherin (AU) expression.