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

Russell Gould[‡], Anirikh Chakrabarti, Jeffrey D. Varner and Jonathan Butcher[‡*]

School of Chemical and Biomolecular Engineering and [‡]Department of Biomedical Engineering, Cornell University, Ithaca NY 14853

Running Title: Modeling of TGF-β induced EMT

Character count: *56,659*

[*]Corresponding author:

Jonathan T. Butcher,

Associate Professor, Department of Biomedical Engineering,

304 Weill Hall, Cornell University, Ithaca NY, 14853

Email: jtb47@cornell.edu

Phone: (607) 255 - 3575

Fax: (607) 255 - 7330

Abstract

Epithelial to mesenchymal transition (EMT) is an essential differentiation program during tissue morphogenesis and remodeling. EMT is induced by soluble transforming growth factor β (TGF-β) 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-β 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-β 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-β 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-β 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-β 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 Thiery (2003). 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 Stahl & Felsen (2001). EMT is the fundamental initiator of developmental processes such as embryonic gastrulation and valvulogenesis Eisenberg & Markwald (1995) (also Kalluri J Clin Invest 2009, Thiery Cell 2009). Transforming growth factor β (TGF-β) family members are important inducers of both de-

velopmental and pathological EMT Xu *et al.* (2009), Zavadil & Böttinger (2005). 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-β isoforms induce EMT across a variety of cell lines Medici *et al.* (2006, 2008). 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 β-catenin, NF-κB, 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 Hardy *et al.* (2010), Huber *et al.* (2004), Jiang *et al.* (2007), Kim *et al.* (2002). 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 Ahmed & Nawshad (2007). Previous computational models of TGF-β induced differentiation focused on single biological factors or EMT in single cells. For example, Chung *et al.*, constructed a model of TGF-β receptor activation and Smad signaling using ordinary differential equations and mass-action kinetics. Their model suggested that a reduction of functional TGF-β receptors in cancer cells may lead to an attenuated Smad2 signal Chung *et al.* (2009). Similarly, Vilar *et al.* suggested that specific changes in receptor trafficking patterns could lead to phenotypes that favor tumorigenesis Vilar *et al.* (2006). 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 Cirit & Haugh (2012) or logical model formulations Morris *et al.* (2011) 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 Bailey (2001). 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" Machta *et al.* (2013). 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 Song *et al.* (2010). Furthermore, we proposed that the sloppy behavior of biological networks may also be seen as a source of cell-to-cell Lequieu *et al.* (2011) or even patient-to-patient heterogeneity Luan *et al.* (2010). Recently, Bayesian parameter identification techniques have also been used to explore cell-to-cell heterogeneity Hasenauer *et al.* (2011), Kalita *et al.* (2011), where a population of cells could be viewed as a dynamic ensemble of context-specific biochemical networks Creixell *et al.* (2012).

In this study, we developed a family of mechanistic models describing the induction of EMT by TGF-β 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-β 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-β1/2 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-β1/2 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-β 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-β induced EMT The EMT model architecture, based upon curated molecular connectivity, described the expression of 80 genes following exposure to TGF-β 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 Chen *et al.* (2009), Schoeberl *et al.* (2002), Tasseff *et al.* (2011). 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 Gadkar *et al.* (2005). 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-β isoforms Medici *et al.* (2008). 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 Lian *et al.* (2011). 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-β exposure. We subdivided the response to TGF-β exposure into two phases. First, TGF-β1/2 signaling initiated a program which downregulated E-cadeherin expression in a MAPK dependent manner while simultaneously upregulating TGF-β3 expression. Second, TGF$-\beta$3 secretion initiated an autocrine feedback which upregulated the expression of mesenchymal markers such as Vimentin and key upstream transcription factors such as LEF-1 in a SMAD dependent manner. Each phase involved the hierarchal expression and/or post-translational modification of several key transcription factors. During the first phase, stimulation with TGF-β1/2 (10 a.u.) activated both the SMAD and MAPK pathways. MAPK activation resulted in the phosphorylation of the transcription factor activator protein 1 (AP-1), which in-turn upregulated the expression of Snail, a well established transcriptional repressor (Fig. 2A). Snail expression was MAPK-dependent; the MEK inhibitor U0126 blocked AP-1 activation and Snail expression following TGF-β1/2 exposure (Fig. 2A, Lane 3). Similar results were obtained for Slug expression, confirming initial activation through the MAPK pathway (data not shown). Overexpression of either Snail or Slug upregulated TGF-β3 expression (Fig. 2C) while simultaneously downregulating E-cadeherin expression (Fig. 2F). During the second phase, TGF$-\beta$3 secretion and the subsequent autocrine signaling resulted in the upregulation of mesenchymal marker expression. The TGF$-\beta$3 induced gene expres-

sion program involves a complex hierarchy of transcriptional and post-translational regulatory events. Absence of E-cadherin indirectly promoted TGF$-\beta$3 expression through the β-catenin/TCF4 complex following Snail or Slug expression (Fig. 2C, Lane 2 or 3). Conversely, over-expression of E-cadherin inhibited the TGF$-\beta$3 autocrine production by sequestering cytosolic β-catenin, thereby blocking EMT (Fig. 2C, Lane 4 or 5). TGF$-\beta$3 signaled through through the Smad pathway to regulate LEF-1 expression and downstream target EMT genes (Fig. 2G). TGF$-\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.* Medici *et al.* (2008) 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-β induced EMT required an additional regulatory protein. This protein, which we called hypothetical regulator 1 (YREG1), was required to mediate between SNAIL/SLUG transcriptional activity and the upregulation of LEF-1 expression following TGF-β1/2 exposure. SNAIL/SLUG are well known transcriptional repressors Dhasarathy

et al. (2011), Hemavathy *et al.* (2000a,b), although there are a few studies which suggest that at least SNAIL can also act as a transcriptional activator Guaita *et al.* (2002). In the model, we assumed the expression of SNAIL/SLUG was likely regulated by AP1/SP1 Jackstadt *et al.* (2013). Thus, upon receiving a TGF-β1/2 signal, the model predicted enhanced SNAIL/SLUG expression, consistent with experimental observations. TGF-β1/2 stimulation also induces LEF-1 expression. However, literature evidence suggested that LEF-1 expression was not strongly dependent upon AP1/SP1 activity Eastman & Grosschedl (1999). Thus, either SNAIL/SLUG are acting as inducers (contrary to substantial biochemical evidence) or, they are repressing the expression of an intermediate repressor. Given the biochemical evidence supporting SNAIL/SLUG as repressors, we created YREG1 a hypothetical intermediate repressor whose expression is downregulated by SNAIL/SLUG. The literature data therefore suggested that YREG1 had two transcriptional targets, LEF-1 and TGF-β3. By adding this regulator, our simulations became consistent with training and literature data. Medici et al. suggested a similar idea where feedback between β-catenin and LEF-1 was likely, although this feedback had yet to be identified Medici *et al.* (2008). Low levels of YREG1 expression were used in all simulations to regulate the formation of the β-catenin-LEF-1 complex. To test the potency of YREG1, we conducted knockdown and over-expression simulations following the addition of TGF-β1/2 (Fig. S8). In the absence of YREG1, most of the population failed to consistently respond to TGF-β1/2 exposure compared to the wild-type (Fig. S8A). Conversely, YREG1 overexpression revealed an exclusively epithelial phenotype following TGF-β1/2 stimulation (Fig. S8B). Overexpression of YREG1 repressed LEF-1 and TGF-β3 expression, thereby not allowing free β-catenin to form the β-catenin-LEF-1 complex which promotes mesenchymal gene expression, or SMAD activity following from autocrine TGF-β3 signaling. Likewise, the abundance of the pSmad2/4-LEF-1 complex was also reduced in cells overexpression YREG1, which blocked the repression of E-cadherin. Taken together, we found that low YREG1 expression was necessary for stabilizing EMT, while elevated

YREG1 levels limited the extent of EMT induction.

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$1/2 exposure, an often neglected but important emergent feature of developmental and pathological programs is population heterogeneity Park *et al.* (2010). We (and others) previously hypothesized that deterministic model ensembles can interrogate population behavior, at least at a course grained level Lequieu *et al.* (2011). We tested this hypothesis by analyzing the response of the population of EMT models to extracellular cues and then comparing this response to flow cytometry studies. We used robustness coefficients to quantify the response of the individual members of the ensemble to TGF$-\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 Lequieu *et al.* (2011), Song *et al.* (2010), Tasseff *et al.* (2010, 2011). Robustness coefficients quantify shifts in molecular marker abundance resulting from molecular of 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 ~ 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$1/2 exposure. Models near P1 had evaluated nuclear localized phosphorylated Sp1, relative to non-induced cells (and models near P2). Elevated Sp1 activity decreased E-cadherin expression through Slug-mediated inhibition, which in turn increased Vimentin expression through TGF$-\beta$3 autocrine signaling and the liberation of β-catenin. Near P2, Sp1 transcriptional activity was lower than P1, leading to only modestly increased Vimentin expression and E-cadherin repression following TGF$-\beta$1/2 stimulation. Near subpopulation three (P3), reduced levels of nuclear phosphorylated AP-1, Sp1, and NFAT (resulting from the loss of ERK kinase activity) were responsible for Vimentin *repression* relative to the control. However, the most biologically interesting behavior was exhibited by subpopulation four (P4). Models near P4 had elevated Sp1 and NFAT transcriptional activity, which increased *both* Vimentin and E-cadherin expression. Analysis of these hypothetical cells suggested they had *abnormal* signaling; deregulated NFAT expression and nuclear localization promoted E-cadherin expression while TGF$-\beta$1/2 induced Sp1 action promoted Vimentin expression. Analysis of the connectivity and information flow through the signaling architecture suggested that Sp1 and NFAT action could be manipulated *independently* by simultaneous TGF$-\beta$1/2 and VEGF-A stimulation (Fig. S1).

To test this hypothesis, we simulated the response of the network to TGF$-\beta$1/2 and VEGF-A treatment with and without NFATc inhibitors (Fig. 4). As expected, stimulation with VEGF-A (50 a.u.) maintained an epithelial population (Q4-43.6%), while TGF$-\beta$1/2 (10 a.u.) exposure shifted the population from an epithelial (Q4-5.5%) to a mesenchymal (Q1-45.6%) phenotype (Fig. 4A and Fig. 4B). On the other hand, combined stimulation with TGF$-\beta$1/2 (10 a.u.) and VEGF-A (50 a.u.) increased both E-cadherin and Vimentin expression (Q2-45.3%), resulting in a hybrid phenotype with both epithelial and mesenchymal characteristics (Fig. 4C). To better understand this hybrid response, we

quantified the simulated protein levels for E-cadherin, Vimentin, phosphorylated nuclear Sp1, nuclear NFATc1, α-smooth muscle actin (α-SMA) and Slug as a function of condition (Fig. S2A-C). Vimentin expression was correlated with high levels of nuclear phosphorylated Sp1, following TGF$-\beta$1/2 exposure. Conversely, elevated E-cadherin expression depended upon the activity of NFAT transcription factors downstream of VEGF-A stimulation. To further isolate the role of NFAT on this hybrid state, we simulated the inhibition of NFAT transcriptional activity across all conditions (all else being equal). NFAT inhibition in combination with VEGF-A treatment blocked all E-cadherin positive sets (Fig. 4D). Likewise, TGF$-\beta$1/2 treatment in combination with NFATc inhibition also resulted in the loss of E-cadherin expression (Fig. 4E). Lastly, NFATc inhibition in combination with simultaneous TGF$-\beta$1/2 and VEGF-A exposure repressed nearly all E-cadherin expression, shifting nearly the entire population towards a mesenchymal phenotype (Fig. 4F). Taken together, high levels of nuclear localized phosphorylated Sp1 correlated with Vimentin expression, while NFATc transcriptional activity was predicted to be critical for maintaining E-cadherin expression.

Combined TGF$-\beta$2 and VEGF-A exposure drives heterogeneity in MCF10A and DLD1 cells The EMT model simulations suggested the transcriptional activity of NFATc and Sp1 could be independently tuned to generate a hybrid cell population with both epithelial and mesenchymal characteristics. To test this hypothesis, we exposed either quiescent epithelial (MCFA10, (Fig. 5)) or transformed epithelial cells (DLD1, (Fig. S3)) to combinations of TGF$-\beta$1/2 and/or VEGF-A. As expected, treatment with TGF$-\beta$1/2 (10ng/ml) increased Slug and Vimentin expression, while repressing E-cadherin expression both at the transcript and protein levels in MCF10A (Fig. 5A-B) and DLD1 cells (Fig. S4C, Fig S3 D,E). Both MCF10A (Fig. 5C) and DLD1 cells (Fig. S3E,G) transitioned from quiescent cobblestone morphology to spread spindle shapes, consistent with EMT. As predicted, we found increased nuclear localization of phosphorylated Sp1 following TGF$-\beta$1/2 stimulation in both MCF10A (Fig. 5B,C) and DLD1 cells (Fig. S3E,F). Con-

sistent 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$1/2 in the presence or absence of VIVIT, a soluble peptide inhibitor of NFATc transcriptional activity Aramburu *et al.* (1999). Treatment with VEGF-A (50ng/ml) and VIVIT (10μM) in MCF10A cells significantly reduced E-cadherin expression compared to VEGF-A alone (Fig 5D,E). Co-treatment with VIVIT and TGF$-\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.

Ductal branching during acini formation is dependent upon phenotype heterogeneity in MCF10A and DLD1 cells We finally employed established three-dimensional (3D) *in vitro* models of invasion, migration, compaction, and tubulogenesis Dhimolea *et al.* (2010) 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$1/2 (Fig. 6B-D). The responses of DLD1 cells followed a similar trend to MCF10A, although the magnitudes of migration, invasion, and compaction were less. Cell circularity within 3D gels strongly and negatively correlated with both invasion and compaction regardless of treatment (Fig. 6E). Circularity refers to the morphology of the cells. In general, a quiescent epithelial cells assumes a circular morphology in culture, while an active mesenchymal cell is highly elongated. The circularity index, a common means of quantifying cell morphology, relates cell area to perimeter. A perfect circle has a circularity index equal to 1.0, while a straight line has a circularity index equal to 0.0, see Butcher et al. Butcher *et al.* (2004). TGF$-\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$1/2 and VEGF-A. No tubular structures were identified within the DLD1 constructs during the 7 day tubulogenesis endpoints, supporting that MCF10A and DLD1 cells have some cell-type specific EMT sensitivity despite their underlying competency for acquiring a heterogeneous phenotype. This suggests that initial EMT sensitivity of a cell influences downstream functional response from TGF-b and VEGFA stimulation. Together, these results establish that VEGF-A and TGF$-\beta$1/2 ligand concentrations potentiate between acini and ductal branch formation in 3D culture, and are dependent upon NFATc activity.

Discussion

In this study, we developed a family of mechanistic models describing the induction of EMT by TGF-β 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 Medici *et al.* (2008). 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-β 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-β and VEGF-A. We tested this hypothesis in MCF10A and DLD1 cells stimulated with combinations of TGF-β and VEGF-A. As expected, in the presence of TGF-β or VEGF-A alone, MCF10A and DLD1 cells were either mesenchymal or epithelial, respectively. However, with both TGF-β 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$-\beta$ Hong *et al.* (2013). VEGF-A mediates pathological angiogenic remodeling of tumors Nagy *et al.* (2007), while TGF$-\beta$ can elicit both protective and oncogenic responses Ferrara (2002), Willis & Borok (2007). 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 Polyak & Weinberg (2009), Strauss *et al.* (2011). 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 Neve *et al.* (2006), Welch-Reardon *et al.* (2014). 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 Zajchowski *et al.* (2001). 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 Mancini & Toker (2009). 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

Singh *et al.* (2010). 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 Debnath *et al.* (2003). It has been postulated that a cellular response reminiscent of partial EMT underlies this process, stimulating further branching and formation of acini Pearson & Hunter (2007). 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 O'Brien *et al.* (2004). 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 Debnath & Brugge (2005). 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 microen-

vironment 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 not commonly observed in human carcinomas Debnath & Brugge (2005). This study focused on the combinatorial effects of two very different ligand families present together in the tumor environment. Additional modeling studies are required to unravel the global response of epithelial cells to the full spectrum of chemical, substrate, and mechanical cues. The simulation strategy presented here is readily adaptable to larger species sets, with the major advantage that experimentally testable hypotheses can be generated regarding how signals get integrated to produce global cellular response. Furthermore, by simulating multiple ensembles of parameter sets, subpopulations across a constellation of phenotypes can be created and mined for common and/or divergent signaling characteristics. This is a significant advantage over forced convergence to a single unique solution and thereby generating a potentially non-physiological homogeneous population.

The deterministic population of EMT models predicted heterogeneous behavior that was qualitatively consistent with experimental studies. There is a diversity of algorithmic approaches to estimate model parameters Moles *et al.* (2003), as well as many strategies to integrate model identification with experimental design Rodriguez-Fernandez *et al.* (2013), Villaverde & Banga (2014). However, despite these advances, the identification of models describing intracellular network behavior remains challenging. There are different schools of thought to deal with this challenge. One school has focused on model reduction. Data-driven approaches Cirit & Haugh (2012), boolean Choi *et al.* (2012) or other logical model formulations Morris *et al.* (2011), Terfve *et al.* (2012) are emerging paradigms that constrain model complexity by the availability of the training and validation data. Other techniques such as constraints based modeling, which is commonly used to model metabolic networks, have also been applied to model transcriptional networks, although primarily in lower eukaryotes and prokaryotes Hyduke & Palsson (2010). These

techniques (and many others, see review Wayman & Varner (2013)) are certainly exciting, with many interesting properties. However, we used the traditional approach of mass action kinetics within an ordinary differential equation framework. The identification problem for the EMT model was massively underdetermined. This is not uncommon for differential equation models, especially those that are highly mechanistic. Of course, we could have discarded mechanism or reduced the model scope to decrease the complexity of the identification problem. However, a central criticism leveled by biologists is that model simplification is often done at the cost of biological reality, or done for reasons of computational expediency Sainani (2012). To avoid this criticism, we systematically identified an ensemble of likely models each consistent with the training data, instead of a single but uncertain best fit model. Previously, we (and others) have suggested that deterministic ensembles could model heterogeneous populations in situations where stochastic computation was not feasible Lequieu *et al.* (2011). Population heterogeneity using deterministic model families has previously been explored for bacterial growth in batch cultures Lee *et al.* (2009). In that case, distributions were generated because the model parameters varied over the ensemble, i.e., extrinsic noise led to population heterogeneity. In this study, parameters controlling physical interactions such as disassociation rates or the rate of assembly or degradation of macromolecular machinery such as ribosomes were widely distributed over the ensemble. Population heterogeneity can also arise from intrinsic thermal fluctuations, which are not captured by a deterministic population of models Swain *et al.* (2002). Thus, deterministic ensembles, provide a coarse-grained or extrinsic-only ability to simulate population diversity. Despite this limitation, our prediction of phenotypic heterogeneity (and the underlying signaling events responsible for the heterogeneity) was consistent with experimental observations. This suggested that deterministic ensembles could simulate disease or developmental processes in which heterogeneity plays an important role, without having to resort to stochastic simulation.

A common criticism of ODE modeling has been the poorly characterized effect of

structural and parametric uncertainty. In this study, parametric uncertainty was addressed by developing an ensemble of probable models instead of a single best-fit but uncertain model using multiobjective optimization. While computationally complex, multiobjective optimization is an important tool to address qualitative conflicts in training data that arise from experimental error or cell line artifacts Handl *et al.* (2007). On the other hand, structural uncertainty is defined as uncertainty in the biological connectivity. The EMT model connectivity was assembled from an extensive literature review. However, several potentially important signaling mechanisms were not included. First, we identified a potential gap in biological knowledge surrounding the regulation of LEF-1 expression, that was filled by the addition of the hypothetical YREG1 transcriptional repressor. The LEF-1 transcription factor is expressed in tissues that undergo EMT during embryogenesis Nawshad & Hay (2003), Vega *et al.* (2004), and has been suggested to promote an invasive phenotype in cancer cells Cano *et al.* (2000), Kim *et al.* (2002). Low levels of YREG1 were important for stabilizing the interaction between LEF-1 and β-catenin, while elevated levels inhibited EMT by downregulating LEF-1 transcriptional activity. Recent evidence has established a complex role of Amino terminal Enhancer of Split (AES) and Groucho/TLE on suppressing LEF-1 activity. AES opposes LEF-1 transcriptional activation while Groucho/TLE binds with LEF-1 for a histone deacetylase repression. In addition, β-catenin directly displaces Groucho/TLE repressors from TCF/LEF-1 in Wnt-mediated transcription activation Arce *et al.* (2009), Grumolato *et al.* (2013). Our model agrees with this newly discovered feedback system, as YREG1 regulates LEF-1 activity leading to EMT stabilization. Secondly, we should revisit the role of GSK-3β. GSK-3β is an important regulator which controls the abundance of both Snail and β-catenin through the ubiquitin-proteasome pathway Larue & Bellacosa (2005), Zhou *et al.* (2004). Specifically to our model, expression of Snail increases through 72 hrs. In contrast, experimental data has shown that activity of Snail peaks at 24 hrs which may be controlled by the GSK-3β complex Medici *et al.* (2006). Recent evidence has also suggested an essential role of NF-κB

in epithelial transformation. NF-κB may influence Snail expression through the AKT pathway and directly stabilize Snail activity Wu *et al.* (2009). This is particularly important for integrating inflammation pathways, such as interleukin-6 (IL-6) and tumor necrosis factor-α (TNF-α), which have been linked to EMT in pathological conditions Sullivan *et al.* (2009). Other pathways such as Notch have also been shown to act synergistically with TGF-β to express Slug in the developing embryo Niessen *et al.* (2008). Lastly, while we have modeled classical protein signaling, we have 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-β exposure Bullock *et al.* (2012). Addressing missing structural components like these, could generate more insight into TGF-β signaling and its role in phenotypic transformation.

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).

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 Jensen *et al.* (2009), Linding *et al.* (2007). 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 though which MDCKII, DLD1 colon carcinoma, and A375 melanoma cells transition towards a mesenchymal phenotype Medici *et al.* (2008). Sequential activation of MAPK and Smad pathways were initiated upon addition of TGFβ1/2. Briefly, TGFβ2 signals through the RAS-RAF-MEK-ERK pathway to up-regulate Snail and Slug expression Medici *et al.* (2006). Snail, a known repressor of junctional proteins, inhibits the expression of E-cadherin Cano *et al.* (2000). This initial repression of E-cadherin leads to a release of β-catenin from the cell membrane. Cystolic β-catenin can then translocate to the nucleus and form transcriptional complexes with TCF-4 to drive TGFβ3 expression Medici *et al.* (2008). TGFβ3 signals to the cells interior by binding to type II receptors, which form heterodimers with type I receptors (ALK5) Derynck & Zhang (2003). This activates the receptors serine/threonine kinase activity to phosphorylate and activate the receptor Smads 2/3 Massagué *et al.* (2005). 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 Nawshad *et al.* (2007). LEF-1 also binds with β-catenin to upregulate mesenchymal proteins such as fibronectin Medici *et al.* (2011). 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 Derynck & Zhang (2003). 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 Gatza *et al.* (2010). Complex assembly activates the serine/threonine kinase activity on the receptor, leading to the recruitment and phosphorylation of Smad partners Massagué *et al.* (2005). 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*) Nawshad *et al.* (2007) 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 Cano *et al.* (2000). However, this transition is not solely driven by transcriptional events. At the protein level, the repression of E-cadherin leads to a release of β-catenin from cell membrane. Cystolic β-catenin then translocates to the nucleus and forms transcriptionally-active complexes with immunoglobulin transcription factor 2 (TCF-4) to drive TGF$-\beta$3 expression Medici *et al.*

(2008). 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:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \cdot \mathbf{r}\left(\mathbf{x}, \mathbf{k}\right) \qquad \mathbf{x}\left(t_o\right) = \mathbf{x}_o \tag{1}$$

The quantity $\mathbf{x}$ denotes the vector describing the abundance of protein, mRNA, and other species in the model (995×1). The stoichiometric matrix $\mathbf{S}$ encodes the signaling architecture considered in the model (995×1700). 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 σ_{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 $\mathbf{r}\left(\mathbf{x}, \mathbf{k}\right)$ denotes the vector of interactions rates (1700×1). 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:

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

The set $\{\mathbf{R}_q\}$ denotes reactants for reaction q, while σ_{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 Hindmarsh *et al.* (2005) on an Apple workstation (Apple, Cupertino, CA) as previously described Tasseff *et al.* (2011).

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.* Medici *et al.* (2008). 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 Song *et al.* (2010). Cross-validation was used to calculate both training and prediction error during the parameter estimation procedure Kohavi (1995). 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 Lequieu *et al.* (2011), Tasseff *et al.* (2011). Robustness coefficients denoted by $\alpha\left(i,j,t_o,t_f\right)$:

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

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 $\alpha\left(i,j,t_o,t_f\right)>1$, indicates increased marker abundance, while $\alpha\left(i,j,t_o,t_f\right)<1$ indicates decreased marker abundance following perturbation j. If $\alpha\left(i,j,t_o,t_f\right)\sim 1$ the jth perturbation does not influence the abundance of marker i. Robustness coefficients were calculated for each member of the ensemble (N $\simeq$ 1100).

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

RT-PCR RNA extractions were performed using a Qiagen total RNA purification kit (Qiagen, Valencia, CA) and RNA was reverse transcribed to cDNA using the SuperScript III RT-PCR kit with oligo(dT) primer (Invitrogen). Sufficient quality RNA was determined by an absorbance ratio A260/A280 of 1.8-2.1, while the quantity of RNA was determined by measuring the absorbance at 260nm (A260). Real-time PCR experiments were conducted using the SYBR Green PCR system (Biorad, Hercules, CA) on a Biorad CFX96 cycler, with 40 cycles per sample. Cycling temperatures were as follows: denaturing, 95C; annealing, 60C; and extension, 70C. Primers were designed to detect GAPDH, E-cadherin, vimentin, Slug, Sp1, and NFATc1 in cDNA clones: Sp1 (F-TTG AAA AAG GAG TTG GTG GC, R-TGC TGG TTC TGT AAG TTG GG, Accession NG030361.1), NFATc1 (F-GCA TCA CAG GGA AGA CCG TGT C, R-GAA GTT CAA TGT CGG AGT TTC TGA G, Accession NG029226.1). GAPDH, E-cadherin, vimentin, and Slug primers were taken from previously published literature Medici *et al.* (2008).

Antibody Staining Samples were fixed in 4% PFA overnight at 4C. Samples were then washed for 15 minutes on a rocker 3 times with PBS, permeabilized with 0.2% Triton-X 100 (VWR International, Radnor, PA) for 10 minutes, and washed another 3 times with PBS. Samples were incubated overnight at 4C in a 1% BSA (Rockland Immunochemicals, Inc., Gilbertsville, PA) blocking solution followed by another 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 (20μL; 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. 250μL of gel was inoculated into culture wells, which solidified after 60min. Treatments were then added within 800μL 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 exam-

ple, using image segmentation, Chang et al. Chang *et al.* (2007) examined the elongation of the MCF10A acini at 6, 12, and 96 hours after a particular treatment. Polizzotti et al. Polizzotti *et al.* (2012) 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.

Statistics 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 $p \prec 0.05$.

References

Abramoff M, Magelhaes P, Ram S (2004) Image Processing with ImageJ. *Biophotonics International,* **11**: 36–42

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

Aramburu J, Yaffe MB, López-Rodríguez C, Cantley LC, Hogan PG, Rao A (1999) Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. *Science* **285**: 2129–33

Arce L, Pate KT, Waterman ML (2009) Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. *BMC Cancer* **9**: 159

Bailey JE (2001) Complex biology with no parameters. *Nat Biotechnol* **19**: 503–4

Bonni S, Wang HR, Causing CG, Kavsak P, Stroschein SL, Luo K, Wrana JL (2001) TGF-beta induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation. *Nat Cell Biol* **3**: 587–595

Bullock MD, Sayan AE, Packham GK, Mirnezami AH (2012) MicroRNAs: critical regulators of epithelial to mesenchymal (EMT) and mesenchymal to epithelial transition (MET) in cancer progression. *Biol Cell* **104**: 3–12

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

Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F, Nieto MA (2000) The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol* **2**: 76–83

Chang H, Park C, Parvin B (2007) Quantitative Representation of Three-dimensional Cell Culture Models. In *Proceedings of the 2007 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington, DC, USA, April 12-16, 2007*. pp. 89–92

742 Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK
743 (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action
744 model trained against dynamic data. *Mol Syst Biol* **5**: 239
745 Choi J, Park SY, Joo CK (2007) Transforming growth factor-beta1 represses E-cadherin
746 production via slug expression in lens epithelial cells. *Invest Ophthalmol Vis Sci* **48**:
747 2708–2718
748 Choi M, Shi J, Jung SH, Chen X, Cho KH (2012) Attractor landscape analysis reveals
749 feedback loops in the p53 network that control the cellular response to DNA damage.
750 *Sci Signal* **5**: ra83
751 Chung SW, Miles FL, Sikes RA, Cooper CR, Farach-Carson MC, Ogunnaike BA (2009)
752 Quantitative modeling and analysis of the transforming growth factor beta signaling
753 pathway. *Biophys J* **96**: 1733–50
754 Cirit M, Haugh JM (2012) Data-driven modelling of receptor tyrosine kinase signalling
755 networks quantifies receptor-specific potencies of PI3K- and Ras-dependent ERK acti-
756 vation. *Biochem J* **441**: 77–85
757 Creixell P, Schoof EM, Erler JT, Linding R (2012) Navigating cancer network attractors for
758 tumor-specific therapy. *Nat Biotechnol* **30**: 842–8
759 Davies M, Robinson M, Smith E, Huntley S, Prime S, Paterson I (2005) Induction of an
760 epithelial to mesenchymal transition in human immortal and malignant keratinocytes by
761 TGF-beta1 involves MAPK, Smad and AP-1 signalling pathways. *J Cell Biochem* **95**:
762 918–931
763 Debnath J, Brugge JS (2005) Modelling glandular epithelial cancers in three-dimensional
764 cultures. *Nat Rev Cancer* **5**: 675–88
765 Debnath J, Muthuswamy SK, Brugge JS (2003) Morphogenesis and oncogenesis of MCF-
766 10A mammary epithelial acini grown in three-dimensional basement membrane cul-
767 tures. *Methods* **30**: 256–268
768 Dennler S, Goumans MJ, ten Dijke P (2002) Transforming growth factor beta signal trans-

duction. *J Leukoc Biol* **71**: 731–40

Derynck R, Zhang YE (2003) Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**: 577–84

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

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

Eastman Q, Grosschedl R (1999) Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr Opin Cell Biol* **11**: 233–40

Eisenberg LM, Markwald RR (1995) Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ Res* **77**: 1–6

Fan JM, Ng YY, Hill PA, Nikolic-Paterson DJ, Mu W, Atkins RC, Lan HY (1999) Transforming growth factor-beta regulates tubular epithelial-myofibroblast transdifferentiation in vitro. *Kidney Int* **56**: 1455–1467

Ferrara N (2002) VEGF and the quest for tumour angiogenesis factors. *Nat Rev Cancer* **2**: 795–803

Fonseca C, Fleming P, *et al.* (1993) Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In *Proceedings of the fifth international conference on genetic algorithms*, vol. 423. Citeseer, pp. 416–423

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

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

Gatza CE, Oh SY, Blobe GC (2010) Roles for the type III TGF-beta receptor in human cancer. *Cell Signal* **22**: 1163–74

Grumolato L, Liu G, Haremaki T, Mungamuri SK, Mong P, Akiri G, Lopez-Bergami P, Arita A, Anouar Y, Mlodzik M, Ronai ZA, Brody J, Weinstein DC, Aaronson SA (2013) ?-Catenin-Independent Activation of TCF1/LEF1 in Human Hematopoietic Tumor Cells through Interaction with ATF2 Transcription Factors. *PLoS Genet* **9**: e1003603

Guaita S, Puig I, Franci C, Garrido M, Dominguez D, Batlle E, Sancho E, Dedhar S, De Herreros AG, Baulida J (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

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

Handl J, Kell DB, Knowles J (2007) Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Trans Comput Biol Bioinform* **4**: 279–92

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

Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, Allgöwer F (2011) Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics* **12**: 125

Hemavathy K, Ashraf SI, Ip YT (2000a) Snail/slug family of repressors: slowly going into the fast lane of development and cancer. *Gene* **257**: 1–12

Hemavathy K, Guru SC, Harris J, Chen JD, Ip YT (2000b) Human Slug is a repressor that localizes to sites of active transcription. *Mol Cell Biol* **20**: 5087–95

Hindmarsh A, Brown P, Grant K, Lee S, Serban R, Shumaker D, Woodward C (2005) SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software* **31**: 363–396

Hong JP, Li XM, Li MX, Zheng FL (2013) VEGF suppresses epithelial-mesenchymal tran-

823 sition by inhibiting the expression of Smad3 and miR-192 a Smad3-dependent mi-
824 croRNA. *Int J Mol Med* **31**: 1436–42
825 Huber MA, Azoitei N, Baumann B, Grünert S, Sommer A, Pehamberger H, Kraut N, Beug
826 H, Wirth T (2004) NF-kappaB is essential for epithelial-mesenchymal transition and
827 metastasis in a model of breast cancer progression. *J Clin Invest* **114**: 569–81
828 Hyduke DR, Palsson BØ (2010) Towards genome-scale signalling network reconstruc-
829 tions. *Nat Rev Genet* **11**: 297–307
830 Jackstadt R, Röh S, Neumann J, Jung P, Hoffmann R, Horst D, Berens C, Bornkamm
831 GW, Kirchner T, Menssen A, Hermeking H (2013) AP4 is a mediator of epithelial-
832 mesenchymal transition and metastasis in colorectal cancer. *J Exp Med* **210**: 1331–50
833 Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth
834 A, Simonovic M, Bork P, von Mering C (2009) STRING 8–a global view on proteins and
835 their functional interactions in 630 organisms. *Nucleic Acids Res* **37**: D412–6
836 Jiang YG, Luo Y, He DI, Li X, Zhang LI, Peng T, Li MC, Lin YH (2007) Role of Wnt/beta-
837 catenin signaling pathway in epithelial-mesenchymal transition of human prostate can-
838 cer induced by hypoxia-inducible factor-1alpha. *Int J Urol* **14**: 1034–9
839 Ju W, Ogawa A, Heyer J, Nierhof D, Yu L, Kucherlapati R, Shafritz DA, Böttinger EP
840 (2006) Deletion of Smad2 in mouse liver reveals novel functions in hepatocyte growth
841 and differentiation. *Mol Cell Biol* **26**: 654–67
842 Jungert K, Buck A, von Wichert G, Adler G, König A, Buchholz M, Gress TM, Ellenrieder V
843 (2007) Sp1 is required for transforming growth factor-beta-induced mesenchymal tran-
844 sition and migration in pancreatic cancer cells. *Cancer Res* **67**: 1563–70
845 Kalita MK, Sargsyan K, Tian B, Paulucci-Holthauzen A, Najm HN, Debusschere BJ,
846 Brasier AR (2011) Sources of cell-to-cell variability in canonical nuclear factor-kB (NF-
847 kB) signaling pathway inferred from single cell dynamic images. *J Biol Chem* **286**:
848 37741–57
849 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

Kavsak P, Rasmussen RK, Causing CG, Bonni S, Zhu H, Thomsen GH, Wrana JL (2000) Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol Cell* **6**: 1365–1375

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

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint Conference on artificial intelligence*, vol. 14. Citeseer, pp. 1137–1145

Kretzschmar M, Doody J, Timokhina I, Massagué J (1999) A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras. *Genes Dev* **13**: 804–16

Larue L, Bellacosa A (2005) Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene* **24**: 7443–54

Lee MW, Vassiliadis VS, Park JM (2009) Individual-based and stochastic modeling of cell population dynamics considering substrate dependency. *Biotechnol Bioeng* **103**: 891–9

Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol* **7**: e1002263

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

Linding R, Jensen LJ, Ostheimer GJ, van Vugt MATM, Jørgensen C, Miron IM, Diella F, Colwill K, Taylor L, Elder K, Metalnikov P, Nguyen V, Pasculescu A, Jin J, Park JG, Samson LD, Woodgett JR, Russell RB, Bork P, Yaffe MB, *et al.* (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* **129**: 1415–26

Lopez-Casillas F, Riquelme C, Perez-Kato Y, Ponce-Castaneda MV, Osses N, Esparza-Lopez J, Gonzalez-Nunez G, Cabello-Verrugio C, Mendoza V, Troncoso V, Brandan E (2003) Betaglycan expression is transcriptionally up-regulated during skeletal mus-

cle 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

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

Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. *Science* **342**: 604–7

Mancini M, Toker A (2009) NFAT proteins: emerging roles in cancer progression. *Nat Rev Cancer* **9**: 810–820

Massague J (2003) Integration of Smad and MAPK pathways: a link and a linker revisited. *Genes Dev* **17**: 2993–2997

Massagué J, Seoane J, Wotton D (2005) Smad transcription factors. *Genes Dev* **19**: 2783–810

Masszi A, Di Ciano C, Sirokmány G, Arthur WT, Rotstein OD, Wang J, McCulloch CAG, Rosivall L, Mucsi I, Kapus A (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

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

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

Medici D, Potenta S, Kalluri R (2011) Transforming growth factor-?2 promotes Snail-mediated endothelial-mesenchymal transition through convergence of Smad-dependent and Smad-independent signalling. *Biochem J* **437**: 515–520

Miettinen PJ, Ebner R, Lopez AR, Derynck R (1994) TGF-beta induced transdifferentiation

904 of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *J*
905 *Cell Biol* **127**: 2021–2036
906 Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a
907 comparison of global optimization methods. *Genome Res* **13**: 2467–74
908 Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training
909 signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative
910 analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* **7**: e1001099
911 Nagy JA, Dvorak AM, Dvorak HF (2007) VEGF-A and the induction of pathological angio-
912 genesis. *Annu Rev Pathol* **2**: 251–75
913 Nawshad A, Hay ED (2003) TGFbeta3 signaling activates transcription of the LEF1 gene
914 to induce epithelial mesenchymal transformation during mouse palate development. *J*
915 *Cell Biol* **163**: 1291–301
916 Nawshad A, Lagamba D, Polad A, Hay ED (2005) Transforming growth factor-beta signal-
917 ing during epithelial-mesenchymal transformation: implications for embryogenesis and
918 tumor metastasis. *Cells Tissues Organs* **179**: 11–23
919 Nawshad A, Medici D, Liu CC, Hay ED (2007) TGFbeta3 inhibits E-cadherin gene expres-
920 sion in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription
921 complex. *J Cell Sci* **120**: 1646–53
922 Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP,
923 Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL,
924 Pinkel D, Albertson DG, Waldman FM, *et al.* (2006) A collection of breast cancer cell
925 lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**: 515–27
926 Niessen K, Fu Y, Chang L, Hoodless PA, McFadden D, Karsan A (2008) Slug is a direct
927 Notch target required for initiation of cardiac cushion cellularization. *J Cell Biol* **182**:
928 315–25
929 O'Brien LE, Tang K, Kats ES, Schutz-Geschwender A, Lipschutz JH, Mostov KE (2004)
930 ERK and MMPs sequentially regulate distinct stages of epithelial tubule development.

Dev Cell **7**: 21–32

Oft M, Peli J, Rudaz C, Schwarz H, Beug H, Reichmann E (1996) TGF-beta1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. *Genes Dev* **10**: 2462–2477

Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K (2010) Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. *Clin Cancer Res* **16**: 876–87

Pearson GW, Hunter T (2007) Real-time imaging reveals that noninvasive mammary epithelial acini can contain motile cells. *J Cell Biol* **179**: 1555–67

Peinado H, Portillo F, Cano A (2004) Transcriptional regulation of cadherins during development and carcinogenesis. *Int J Dev Biol* **48**: 365–75

Peinado H, Quintanilla M, Cano A (2003) Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. *J Biol Chem* **278**: 21113–21123

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

Polizzotti L, Basak O, Bjornsson C, Shubert K, Yener B, Plopper G (2012) Novel Image Analysis Approach Quantifies Morphological Characteristics of 3D Breast Culture Acini with Varying Metastatic Potentials. *J Biomed Biotech* **2012**: 1–16

Polyak K, Weinberg RA (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**: 265–273

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

Sainani KL (2012) Meet the Skeptics: Why some doubt biomedical models - and what it takes to win them over. *Biomedical Computation Review* : 12 – 18

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

Singh G, Singh SK, König A, Reutlinger K, Nye MD, Adhikary T, Eilers M, Gress TM, Fernandez-Zapico ME, Ellenrieder V (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

Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using Pareto Optimal Ensemble Techniques (POETs). *Biotechnol J* **5**: 768–80

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

Strauss R, Li ZY, Liu Y, Beyer I, Persson J, Sova P, Möller T, Pesonen S, Hemminki A, Hamerlik P, Drescher C, Urban N, Bartek J, Lieber A (2011) Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLoS One* **6**: e16186

Sullivan NJ, Sasser AK, Axel AE, Vesuna F, Raman V, Ramirez N, Oberyszyn TM, Hall BM (2009) Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. *Oncogene* **28**: 2940–7

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

Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, Varner JD (2010) Analysis of the molecular networks in androgen dependent and independent prostate cancer revealed fragile and robust subsystems. *PLoS One* **5**: e8864

Tasseff 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

985 Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, Morris MK, van Iersel M,
986 Lauffenburger DA, Saez-Rodriguez J (2012) CellNOptR: a flexible toolkit to train protein
987 signaling networks to data using multiple logic formalisms. *BMC Syst Biol* **6**: 133
988 Thiery JP (2003) Epithelial-mesenchymal transitions in development and pathologies.
989 *Curr Opin Cell Biol* **15**: 740–6
990 Tian F, Byfield SD, Parks WT, Stuelten CH, Nemani D, Zhang YE, Roberts AB (2004)
991 Smad-binding defective mutant of transforming growth factor beta type I receptor en-
992 hances tumorigenesis but suppresses metastasis of breast cancer cell lines. *Cancer*
993 *Res* **64**: 4523–4530
994 Tian F, DaCosta Byfield S, Parks WT, Yoo S, Felici A, Tang B, Piek E, Wakefield LM,
995 Roberts AB (2003) Reduction in Smad2/3 signaling enhances tumorigenesis but sup-
996 presses metastasis of breast cancer cell lines. *Cancer Res* **63**: 8284–8292
997 Valcourt U, Kowanetz M, Niimi H, Heldin CH, Moustakas A (2005) TGF-beta and
998 the Smad signaling pathway support transcriptomic reprogramming during epithelial-
999 mesenchymal cell transition. *Mol Biol Cell* **16**: 1987–2002
1000 Vega S, Morales AV, Ocaña OH, Valdés F, Fabregat I, Nieto MA (2004) Snail blocks the
1001 cell cycle and confers resistance to cell death. *Genes Dev* **18**: 1131–43
1002 Vilar JMG, Jansen R, Sander C (2006) Signal processing in the TGF-beta superfamily
1003 ligand-receptor network. *PLoS Comput Biol* **2**: e3
1004 Villaverde AF, Banga JR (2014) Reverse engineering and identification in systems biology:
1005 strategies, perspectives and challenges. *J R Soc Interface* **11**: 20130505
1006 Wayman J, Varner J (2013) Biological systems modeling of metabolic and signaling net-
1007 works. *Curr Opin Chem Eng* **2**: 365 – 372
1008 Welch-Reardon KM, Wu N, Hughes CCW (2014) A Role for Partial Endothelial-
1009 Mesenchymal Transitions in Angiogenesis? *Arterioscler Thromb Vasc Biol*
1010 Willis BC, Borok Z (2007) TGF-beta-induced EMT: mechanisms and implications for fi-
1011 brotic lung disease. *Am J Physiol Lung Cell Mol Physiol* **293**: L525–34

1012 Willis BC, Liebler JM, Luby-Phelps K, Nicholson AG, Crandall ED, du Bois RM, Borok
1013 Z (2005) Induction of epithelial-mesenchymal transition in alveolar epithelial cells by
1014 transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *Am J*
1015 *Pathol* **166**: 1321–1332

1016 Wu Y, Deng J, Rychahou PG, Qiu S, Evers BM, Zhou BP (2009) Stabilization of snail by
1017 NF-kappaB is required for inflammation-induced cell migration and invasion. *Cancer*
1018 *Cell* **15**: 416–28

1019 Xie L, Law BK, Aakre ME, Edgerton M, Shyr Y, Bhowmick NA, Moses HL (2003) Trans-
1020 forming growth factor beta-regulated gene expression in a mouse mammary gland ep-
1021 ithelial cell line. *Breast Cancer Res* **5**: R187–R198

1022 Xie L, Law BK, Chytil AM, Brown KA, Aakre ME, Moses HL (2004) Activation of the Erk
1023 pathway is required for TGF-beta1-induced EMT in vitro. *Neoplasia* **6**: 603–610

1024 Xu J, Lamouille S, Derynck R (2009) TGF-beta-induced epithelial to mesenchymal transi-
1025 tion. *Cell Res* **19**: 156–72

1026 Zajchowski DA, Bartholdi MF, Gong Y, Webster L, Liu HL, Munishkin A, Beauheim C,
1027 Harvey S, Ethier SP, Johnson PH (2001) Identification of gene expression profiles that
1028 predict the aggressive behavior of breast cancer cells. *Cancer Res* **61**: 5168–78

1029 Zavadil J, Bitzer M, Liang D, Yang YC, Massimi A, Kneitz S, Piek E, Bottinger EP (2001)
1030 Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta.
1031 *Proc Natl Acad Sci U S A* **98**: 6686–6691

1032 Zavadil J, Böttinger EP (2005) TGF-beta and epithelial-to-mesenchymal transitions.
1033 *Oncogene* **24**: 5764–74

1034 Zhang YE (2009) Non-Smad pathways in TGF-beta signaling. *Cell Res* **19**: 128–139

1035 Zhou BP, Deng J, Xia W, Xu J, Li YM, Gunduz M, Hung MC (2004) Dual regulation of Snail
1036 by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transi-
1037 tion. *Nat Cell Biol* **6**: 931–40

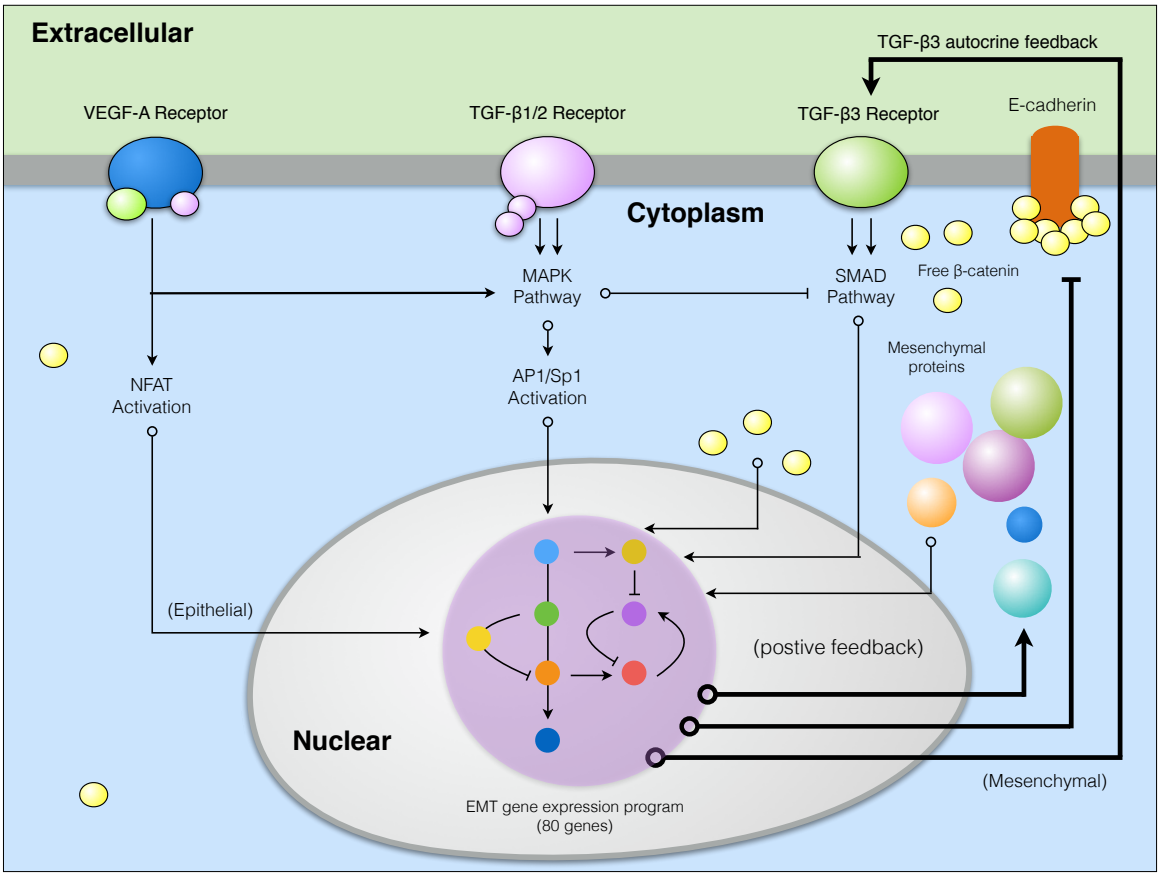

Fig. 1: Model connectivity recreates the core architecture during EMT. The EMT network contains 995 nodes (proteins, mRNA, and genes) interconnected by 1700 interactions. Central to EMT induction, activation of the MAPK cascade occurs through TGF-β1/2 binding which activates the AP-1/Sp1 transcriptional axis. AP-1/Sp1 drives an autocrine response of TGF-β3, 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.

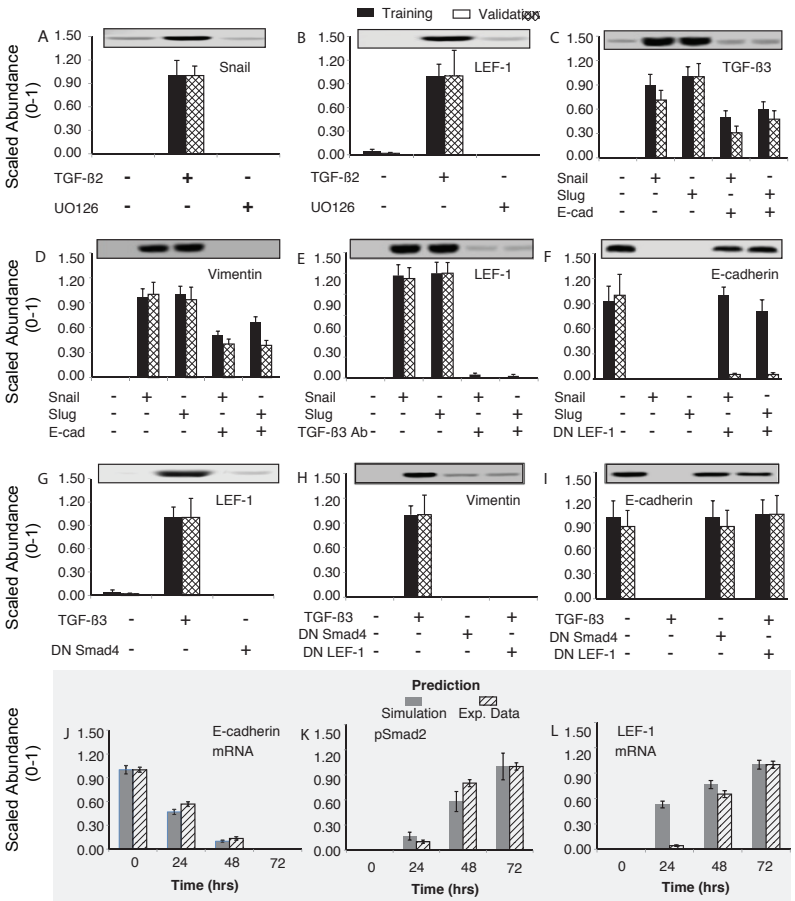

Fig. 2: Training and validation simulations. The population of EMT models qualitatively captured TGFβ-induced EMT signaling. (A-I) The population was generated using POETs and trained using 11 different objective functions (41 data sets) taken from Medici *et al.* Medici *et al.* (2008). The model captured 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.

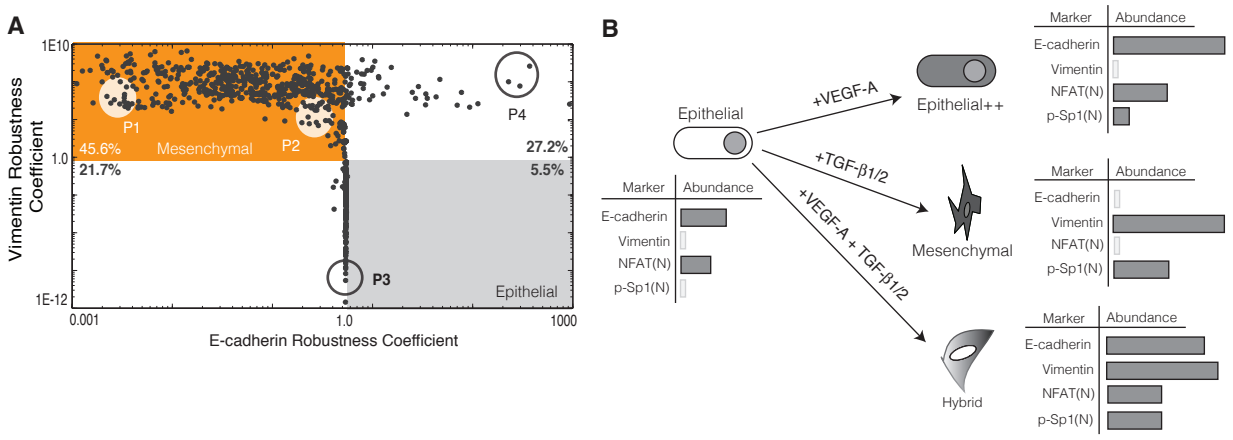

Fig. 3: Simulated TGFβ1/2 exposure promoted phenotype heterogeneity. Robustness coefficients were used to quantify the effect of perturbations after steady state across the population of data sets. Coefficients with values > 1 (< 1) indicated a marker increased (decreased) compared to a base state, while a value of 1 indicated approximately no change following a perturbation. E-cadherin and Vimentin robustness coefficients were used as phenotypic markers. (A) TGFβ1/2 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 TGFβ2 and VEGFA pathway, respectively. (C) We hypothesized that elevated phosphorylated-Sp1 and NFAT levels could drive phenotype heterogeneity through simultaneous TGFβ2 and VEGFA treatment, similar to region three (R3).

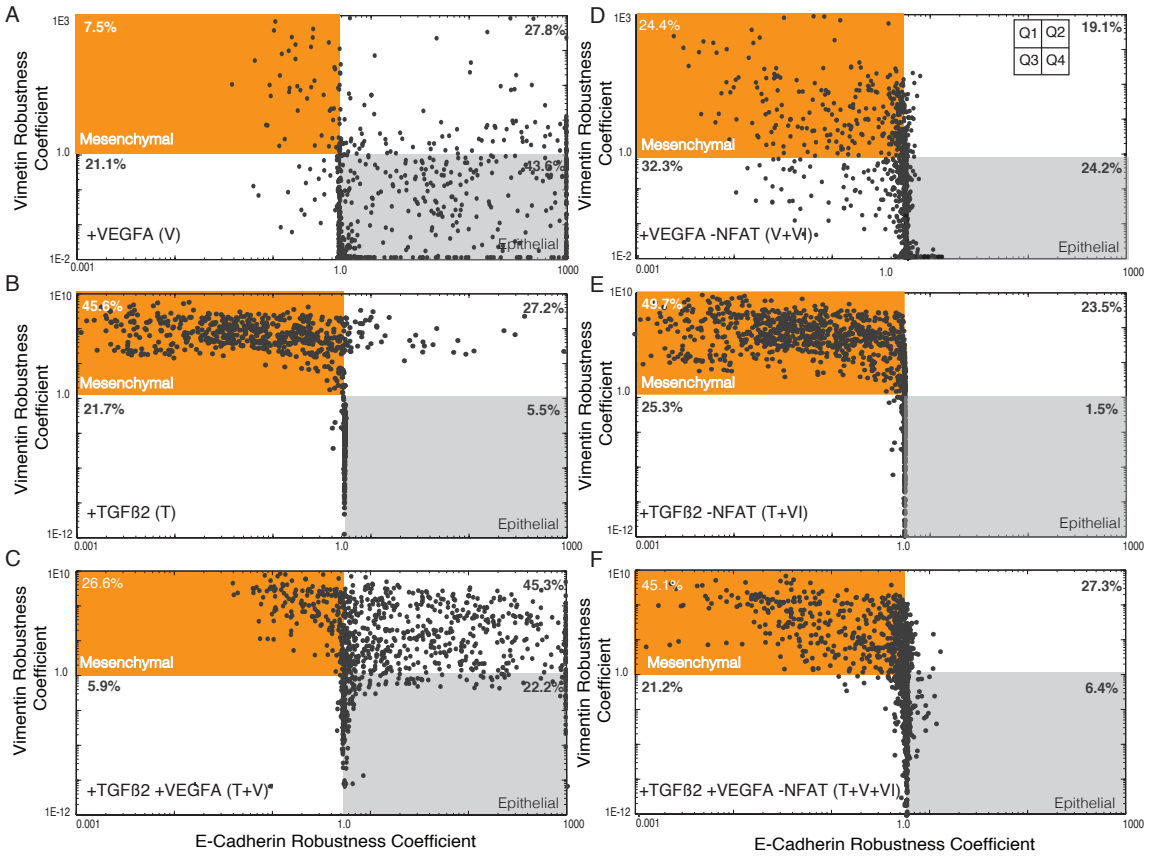

Fig. 4: Simulated response to TGFβ1/2 and VEGF-A exposure with and without axis specific inhibitors. Robustness coefficients were used to quantify the shift in population at 48 hrs. (A-C) VEGF-A (50 a.u.) treatment resulted in a population with enhanced epithelial (Q4) properties. This was contrary to the addition of TGFβ2 (10 a.u.), which shifted the population towards a mesenchymal phenotype (Q1). Interestingly, the combined effects of TGFβ2 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 TGFβ mitigated all ecadherin expression (Q2) confirming their importance for population heterogeneity. Lastly, combination of TGFβ2, 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.

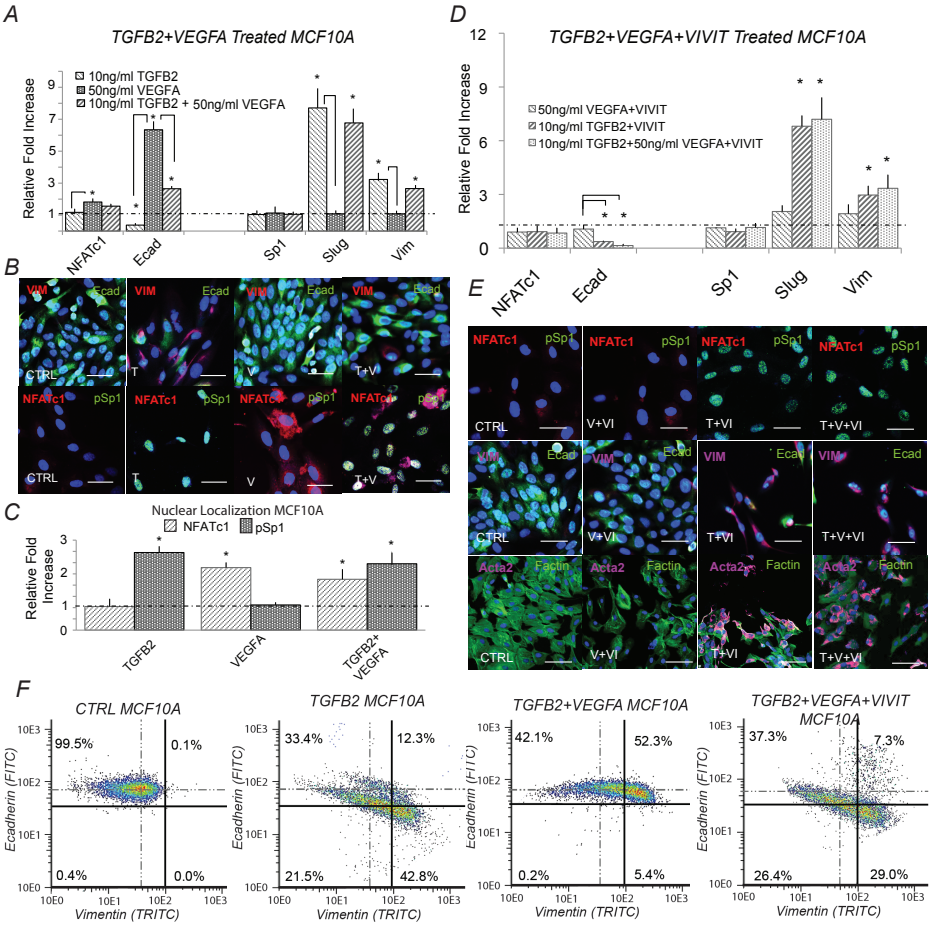

Fig. 5: Simultaneous TGF-β1/2 and VEGF-A treatment induced phenotype heterogeneity and is dependent upon NFAT activity *in-vitro*. (A) In MCF10A, treatment with (10ng/ml) TGFβ2 increased Slug and vimentin, while ecadherin expression was inhibited at both the gene and protein level at 48 hrs. Conversely, VEGFA alone increased both NFATc1 and ecadherin gene expression. Simultaneous TGFβ2 (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 TGFβ2 and VEGFA, respectively. (D) To isolate the effect of NFAT, treatment of VEGFA (50ng/ml) and VIVIT (10μM) reduced ecadherin expression at 48hrs (control-dashed line). Similarly, combined TGFβ2, 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: 50μm. C=Control, T=TGFβ2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ($p \prec 0.05$).

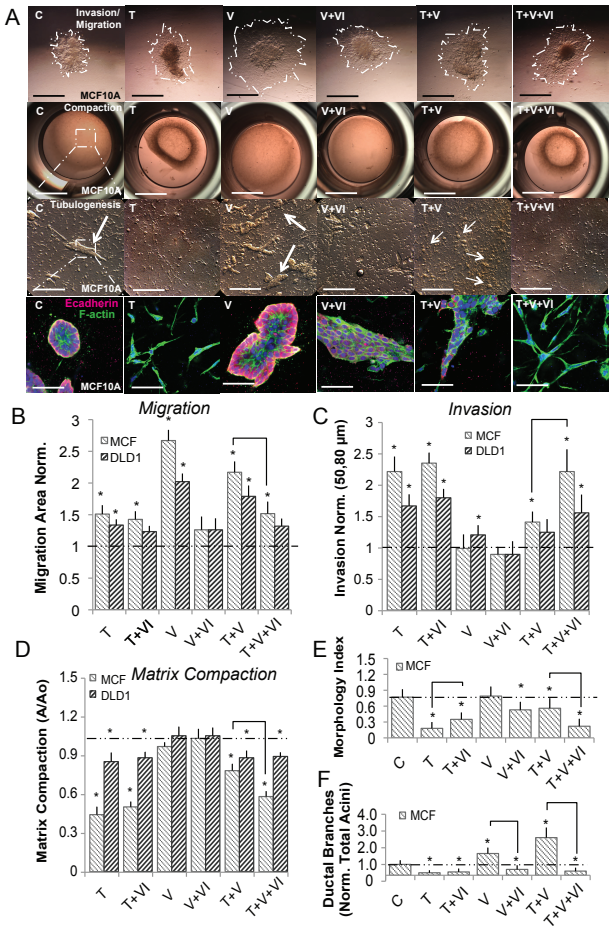

Fig. 6: Ductal branching is dependent upon phenotype heterogeneity within MCF10A in 3-D culture. MCF10A and DLD1 were formed into spheroids overnight and explanted to a collagen gel for 72 hrs. For compaction and tubular assays, cells were embedded into collagen gels for 72 hrs, and the extent of tubulogenesis was measured at 7 days. (A-D) Within MCF10A, TGFβ2 (10ng/ml) enhanced invasion and contractile properties while, VEGFA (50ng/ml) promoted increased migration. TGFβ2 with VEGFA significantly increased migration, while limiting with compaction. VIVIT (10μM) in combination with VEGFA and TGFβ2 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 TGFβ2 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: 500μm, 1000μm, 250μm, and 80μm, respectively. C=Control, T=TGFβ2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ($p \prec 0.05$). 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

EMT model network architecture. TGFβ is a major inducer of EMT in development, fibrosis, and carcinogenesis with different isoforms mediating various effects depending on specific cellular context Nawshad *et al.* (2005). TGFβ1 was first described as an inducer of EMT in normal mammary epithelial cells Miettinen *et al.* (1994) 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 Fan *et al.* (1999), Hales *et al.* (1994), Kasai *et al.* (2005), Willis *et al.* (2005). TGFβ signaling occurs through the Smad pathway in which signals are transduced by transmembrane serine/threonine kinase type I (ALK5) and type II (TGFβRII) receptors. To increase ligand affinity, betaglycan (TGFβRIII) can also interact with TGFβRI,II Gatza *et al.* (2010). Upon TGFβ 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 Massagué *et al.* (2005). 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 TGFβ responsive genes, including alpha-smooth muscle actin, collagen1A2, vimentin, fibronectin, and plasminogen activator inhibitor-1 (PAI-1) by interacting with Smad-binding elements Dennler *et al.* (2002), Derynck & Zhang (2003). To regulate TGFβ signaling, Smurf2 (a ubiquitin E3 ligase) can become activated to mediate proteasome dependent degradation of Smad2 or bind with Smad 7 to target TGFβ receptor for degradation Bonni *et al.* (2001), Kavsak *et al.* (2000).

Cell Type Dependency Interestingly, differential roles for Smad2 and Smad3 in TGFβ 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 Ju *et al.* (2006). 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 Phanish *et al.* (2006). Furthermore, Smad2 signaling has been demonstrated in cancerous lung epithelial A549 cells Kasai *et al.* (2005). Inhibition of Smad3 signaling decreases the metastatic potential of xenografted breast cancer cell lines Tian *et al.* (2004, 2003). Together, these results suggest that the precise Smad pathway activated may depend on the particular cellular context. Regardless, a recent transcriptomic analysis of TGFβ 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 Valcourt *et al.* (2005).

Smad-Independent EMT Induction Pathways. Although less well established than the Smad-dependent pathways in the induction of EMT, there is substantial evidence for TGFβ activation of 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 TGFβ 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 Masszi *et al.* (2003). TGFβ has been shown to elevate ERK activity in culture models of EMT (human keratinocytes, NMuMG mammary epithelial cells, and mouse cortical tubule epithelial cells) Davies *et al.* (2005), Xie *et al.* (2004), Zavadil *et al.* (2001). This ERK activity was required for disassembly of adherens junctions and induction of cell motility Xie *et al.* (2004). Inhibition of ERK blocked these key morphological changes of EMT in mammary gland epithelial cells Xie *et al.* (2003). Recent studies in MDCKII and DLD1 colon carcinoma cells revealed that TGFβ3 can signal through PI3 kinase. Upon activation, PI3 kinase activates molecules such as ILK and AKT, which can phosphory-

late and inactivate GSK3β, a protein that targets both Snail and β-catenin for degradation through the ubiquitin proteosome pathway Medici *et al.* (2008). Lastly, there is increasing evidence for a role of Notch pathways in regulating EMT, particularly during development Nawshad *et al.* (2005). 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 Zavadil & Böttinger (2005).

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 Cano *et al.* (2000), repression of E-cadherin leads to dissolution of adherens junctions. Snail and Slug can be activated by TGFβ via both Smad-dependent and -independent pathways in a cell type dependent fashion in cultured cells Peinado *et al.* (2003). Differential expression of Snail and Slug is observed in TGFβ 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 TGFβ1 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) Peinado *et al.* (2004). Likewise, TGFβ1 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 TGFβ1 induced Slug expression and upstream of MAPK signaling. In addition, the TGFβ1 mediated repression of E-cadherin was significantly inhibited by Slug siRNA Choi *et al.* (2007).

ERK/MAPK Crosstalk The ERK-MAPK pathway has been shown to modify TGFβ signaling at multiple levels. One way is that the activated Ras pathway inhibits the TGFβ induced nuclear accumulation of Smad2/3, as seen in epithelial cells Zhang (2009). 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 TGFβ receptor phosphorylation sites which activate Smad nuclear translocation Massague (2003). 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 TGFβ through attenuation of Smad accumulation in the nucleus Kretzschmar *et al.* (1999). In contrast, Ras signals strongly cooperated with Smads for invasion of human carcinoma cells Oft *et al.* (1996). 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 TGFβ stimulation Jungert *et al.* (2007). Likewise, binding sites at the SP1/Smad3 complex was found to regulate the betaglycan receptors Lopez-Casillas *et al.* (2003). Taken together, multiple levels of cross-talk (both positive and negative) exist within MAPK and TGFβ, 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-β2 (-TGF-β1/2), (b) TGF-β1/2 stimulation (+TGF-β2), and (c) TGF-β1/2 stimulation while blocking the autocrine response of TGF-β3 (+TGF-β1/2 + TGF-β3Ab). Following TGF-β1/2 stimulation, AP1/SP1 rapidly stimu-

lates 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 β-catenin, and the accumulation of free cytosolic β-catenin ensues. Conversely, in the presences of a TGF-β3 neutralizing antibody, membrane bound β-catenin increased after 10 hrs, while low levels of free cytosolic β-catenin accumulated between 1-5 hrs (Fig. S2A,B). Accumulation of free β-catenin is critical for complexing with TCF4 to rapidly produce the autocrine response of TGF-β3 within 1-10 hrs. Following TGF-β3 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. S1C). This timely formation is critical for completion of the EMT process. At steady state, the E-cadherin complex was maintained at a basal epithelial level with no expression of Vimentin (Fig. S1D). During TGF-β1/2 stimulation, both MAPK and Smad act synergistically to repress the E-cadherin complex within 1 hr. This is followed by increase of Vimentin at $\sim$ 1hr, while exponentially increasing around 10 hrs. Both species elevate within 1 hr, however the complex does not significantly form until $\sim$ 10 hrs. When blocking TGF-β3, Snail/Slug downregulates the E-cadherin complex between 5-10 hrs, but cannot complete the transformation due to the loss of Smad2 phosphorylation (Fig. S1D).

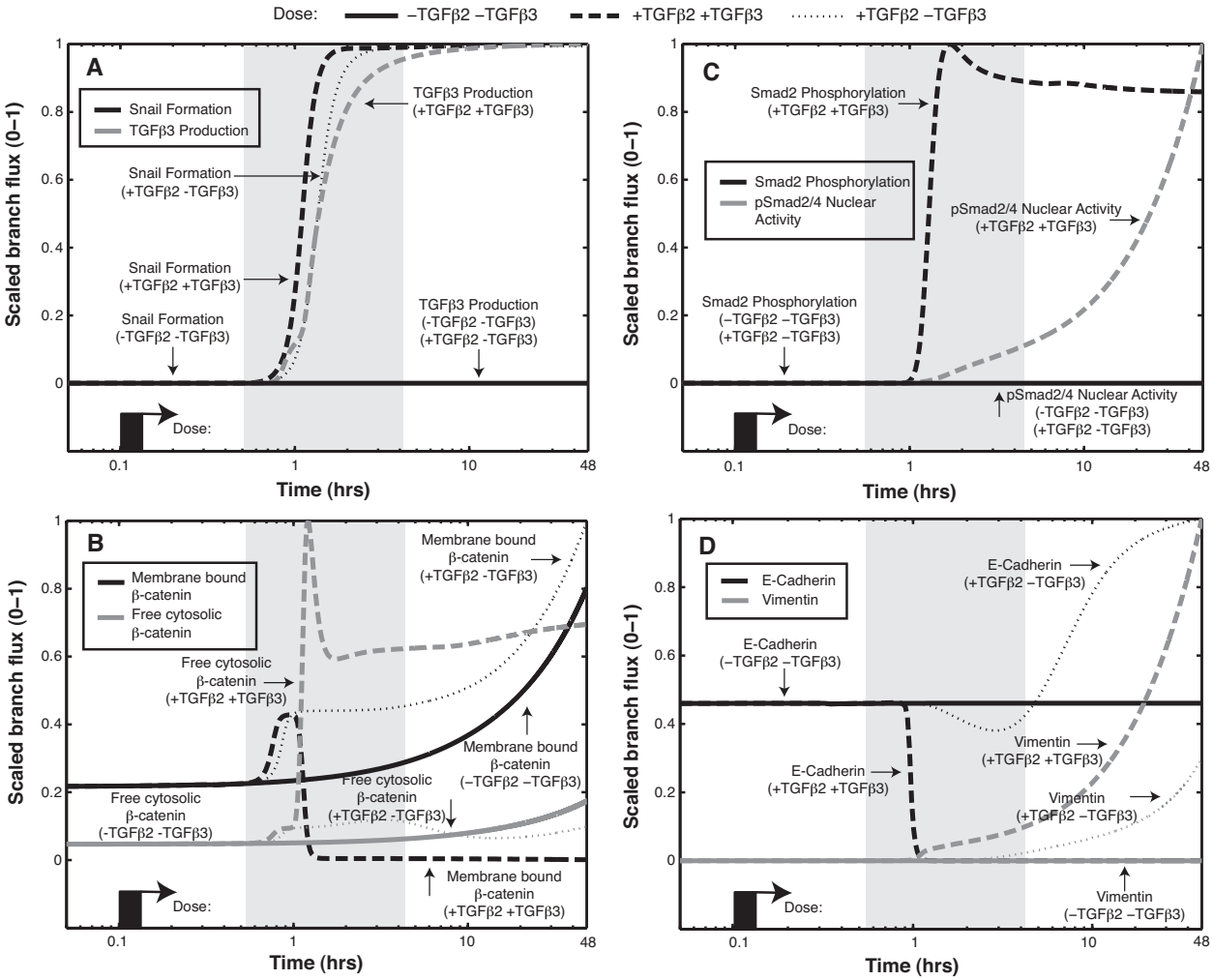

Fig. S1: Signal flow analysis of key species at steady state, TGF-β1/2 stimulation, and blocking the TGF-β3 autocrine response. (A) The MAPK cascade is directly responsible for rapid expression of Snail and downstream TGF-β3 formation (1 hrs). (B) TGF-β1/2 reduces β-catenin, allowing rapid free-cytosolic β-catenin to accumulate (1hr). Blocking TGFβ3 increases membrane bound β-catenin (10hr). (C) TGF-β3 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-β1/2 rapidly reduces the E-cadherin complex, while upregulating Vimentin (5-10 hrs). Blocking TGFβ3 increases E-cadherin (10 hrs) and Vimentin is significantly reduced.

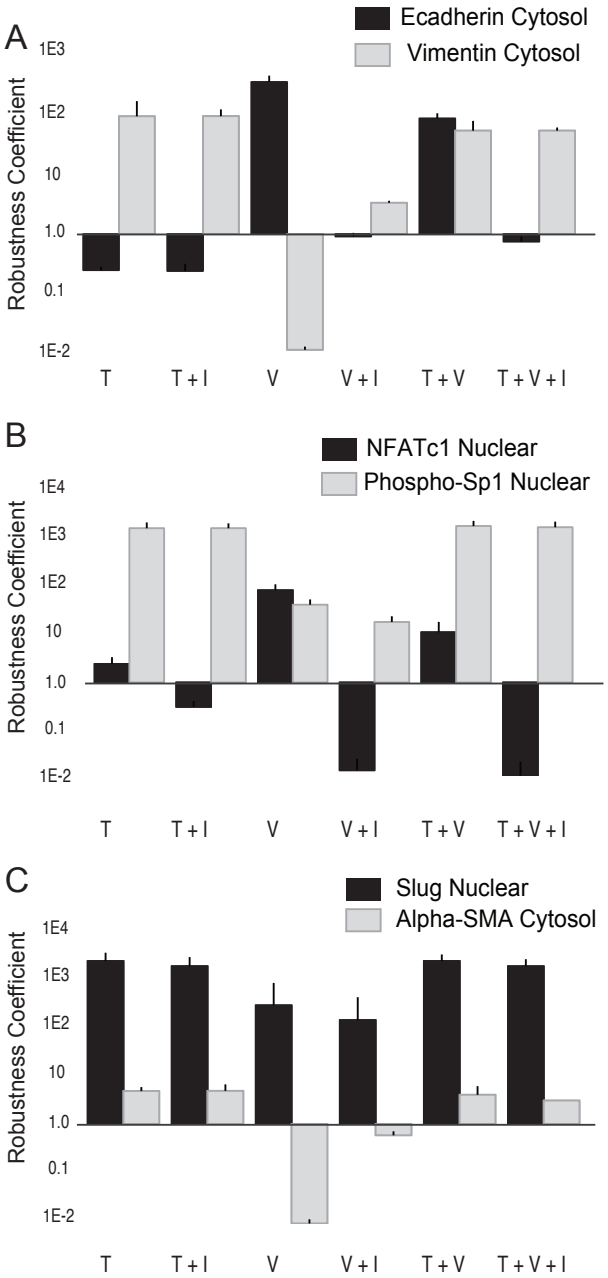

Fig. S2: Robustness analysis for key molecular species at t = 48 hrs for combinations of TGF-β1/2, VEGF-A and NFATc1 inhibitors. Robustness coefficients for the indicated species were calculated for N ~ 1100 ensemble members for 48 hrs following the addition of TGF-β1/2 (T), TGF-β1/2 + NFATc1 inhibitor (T + I), VEGF-A (V), VEGF-A + NFATc1 inhibitor (V + I) and TGF-β1/2 + 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 α-smooth muscle actin (α-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=TGFβ2 , V=VEGFA, VI= NFAT inhibitor (VIVIT).

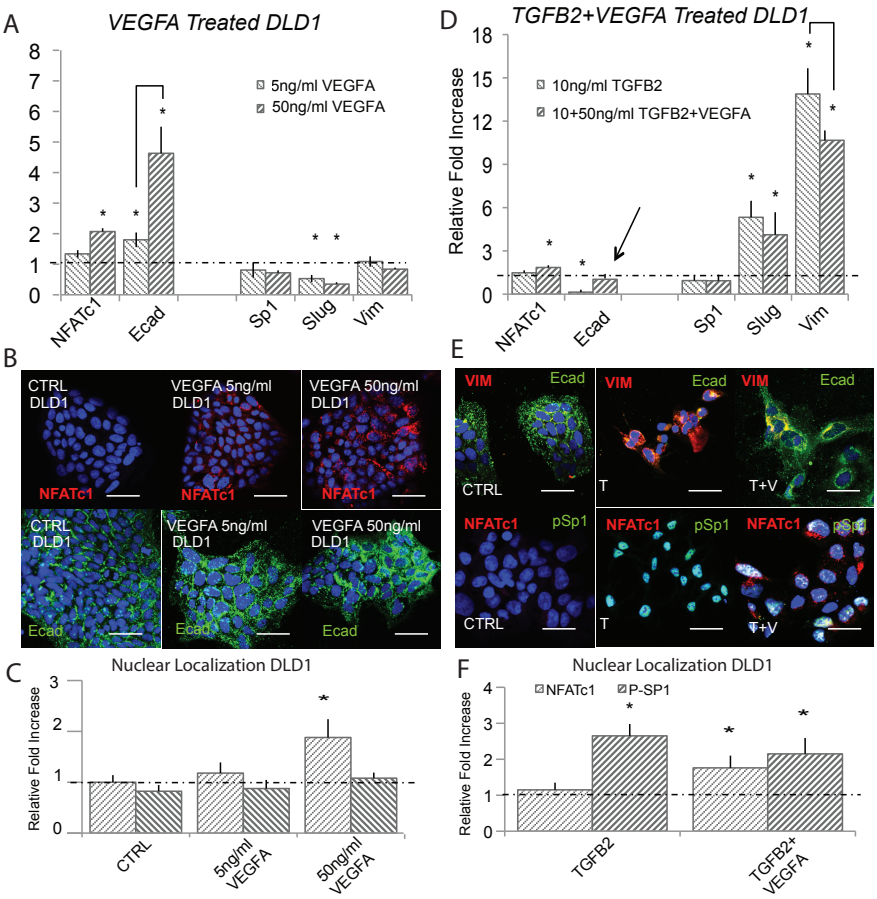

Fig. S3: VEGF-A attenuates TGF-β1/2 to induce phenotype heterogeneity in DLD1. (A) In DLD1, we found that 5ng/ml of VEGFA increased NFATc1 and E-cadherin gene expression via qPCR and 50ng/ml potentiated this effect at 48 hrs. (B - C) These findings were confirmed at the protein level via immunofluorescence, as ecadherin levels and nuclear localization of NFATc1 increased. (D) Treatment with (10ng/ml) TGFβ2 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 TGFβ2 (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: 50μm. C=Control, T=TGFβ2 , V=VEGFA, VI=NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ($p \prec 0.05$).

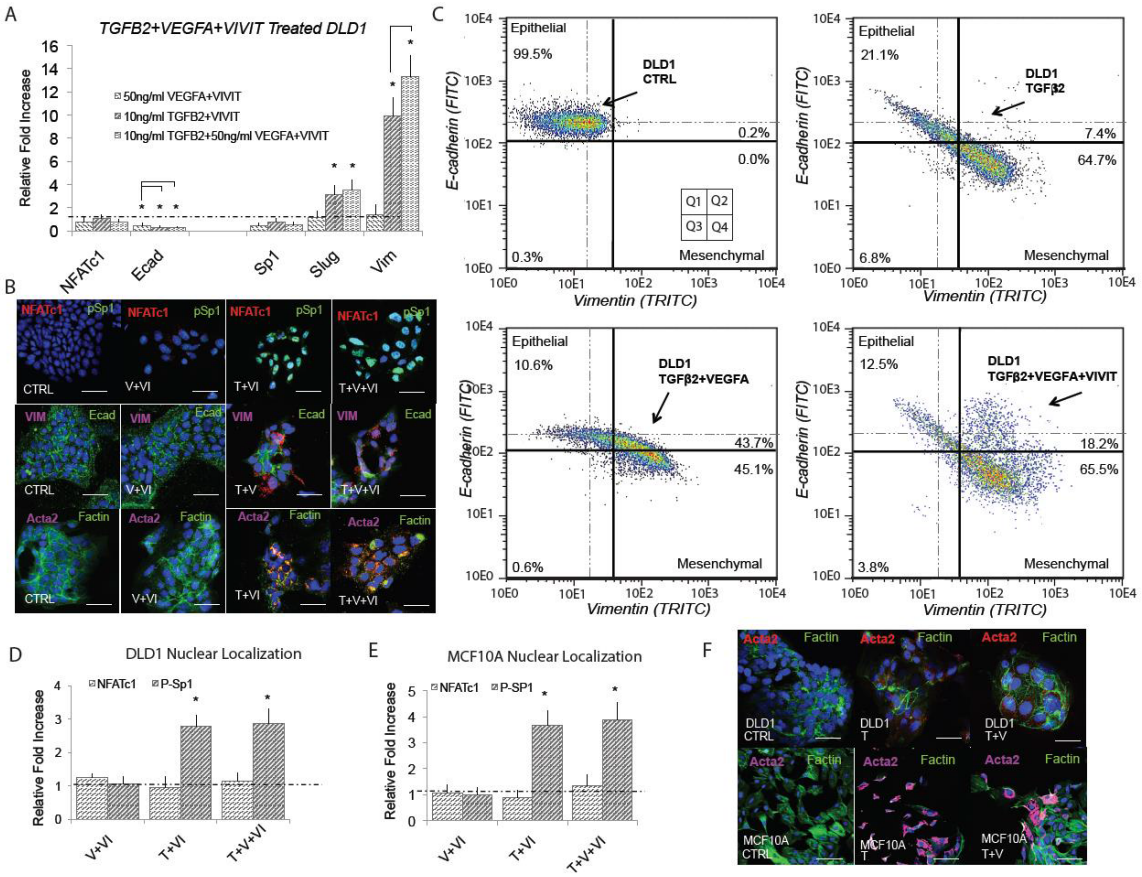

Fig. S4: E-cadherin expression is dependent upon NFAT activity in DLD1. (A) Treatment with VEGFA (50ng/ml) and NFAT inhibitory peptide VIVIT (10μM) resulted in significantly reduced ecadherin expression (qRT-PCR at 48hrs). Addition of TGFβ2 (10ng/ml) and VIVIT resulted in increased Slug and vimentin expression, while inhibiting ecadherin levels. Combined TGFβ2, 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 TGFβ conditions. (C) Quantitative flow cytometry also confirmed this trend. (D,E) TGFβ2, 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 TGFβ conditions. Magnification, 40x. Scale bars: 50μm. C=Control, T=TGFβ2 , V=VEGFA, VI= NFAT inhibitor (VIVIT). Asterisks signify statistical differences from each other according to a one-way ANOVA with Tukey's post hoc ($p \prec 0.05$).

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 TGFβ/EMT models. Cross-validation was used to calculate both training and prediction error during the parameter estimation procedure Kohavi (1995). 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 Song *et al.* (2010). 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:

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

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 Abramoff *et al.* (2004). The scaled measurement for species x at time $i = \{t_1, t_2, .., t_n\}$ in condition j is given by:

$$\hat{\mathcal{M}}_{ij} = \frac{\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}}{\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}} \tag{S2}$$

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) (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.

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 Fonseca *et al.* (1993) 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 $\mathcal{P}\left(\mathbf{k}_{i+1}\right)$:

$$\mathcal{P}(\mathbf{k}_{i+1}) \equiv \exp\left\{-rank\left(\mathbf{k}_{i+1} \mid \mathbf{K}_i\right)/T\right\} \tag{S3}$$

where T is the annealing temperature and $rank\left(\mathbf{k}_{i+1} \mid \mathbf{K}_i\right)$ denotes the Pareto rank for $\mathbf{k}_{i+1}$. The annealing temperature was discretized into 10 quanta between T_o and T_f and

O#	Species (cystolic 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	TGFbeta3	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	"

Fig. S5: Training and prediction values for the 11 TGF-β objective functions versus a random parameter control.

adjusted according to the schedule $T_k = \beta^k T_0$ where β 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 Gadkar *et al.* (2003). From the 15,000 probable EMT models, we selected N = 1093 models with Pareto rank ≤ 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.

Mean Trained vs. Cross Validation = (37% significantly different)

ORI-1	Mean Trained					S.E. Trained				
Smad	0.00	1.00	0.00			Smad	0.00	0.09	0.00	
Shg	0.00	0.00	1.00			Shg	0.00	0.00	0.19	
LEF1	0.00	0.73	1.00			LEF1	0.00	0.22	0.37	
Smad	-	-	+			Smad	-	-	+	
Shg	-	+	-			Shg	-	+	-	
ORI-2	Mean Trained					S.E. Trained				
E-cad	0.92	0.00	0.00	1.00	0.81	E-cad	0.19	0.00	0.00	0.10	0.14
VIM	0.00	0.96	1.00	0.00	0.00	VIM	0.00	0.11	0.11	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-3	Mean Trained					S.E. Trained				
TGFB2	0.00	1.00	0.85			LEF1	0.00	0.13	0.12	
Smad	-	-	+			Smad	-	-	+	
Shg	-	+	-			Shg	-	+	-	
ORI-4	Mean Trained					S.E. Trained				
LEF1	0.00	0.98	1.00	0.03	0.02	LEF1	0.00	0.10	0.11	0.00	0.01
E-cad	1.00	0.00	0.00	0.54	0.35	E-cad	0.20	0.00	0.00	0.16	0.12
VIM	0.00	1.00	0.94	0.00	0.00	VIM	0.00	0.13	0.12	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
TGFB1-Ab	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-5	Mean Trained					S.E. Trained				
B-catenin	0.01	0.99	1.00	0.00	0.25	B-catenin	0.00	0.25	0.26	0.00	0.06
TCF4	0.87	0.23	0.00	0.77	1.00	TCF4	0.27	0.07	0.00	0.26	1.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
E-cad	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-6	Mean Trained					S.E. Trained				
TGFB2	0.00	0.89	1.00	0.49	0.60	TGFB2	0.00	0.16	0.12	0.00	0.09
E-cad	0.98	0.00	0.00	1.00	0.71	E-cad	0.18	0.00	0.00	0.15	0.12
VIM	0.08	0.97	1.00	0.51	0.66	VIM	0.00	0.30	0.10	0.05	0.07
Smad	-	+	-	+	-	Smad	-	-	+	-	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
E-cad	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-7	Mean Trained					S.E. Trained				
Smad	0.00	0.00				Smad	0.00	0.00			
Shg	0.00	0.00				Shg	0.00	0.00			
TGFB2	-	+				Smad	-	+			
ORI-8	Mean Trained					S.E. Trained				
LEF1	0.05	1.00	0.00			LEF1	0.02	0.10	0.00		
TGFB2	-	-	+			Smad	-	-	+		
IN Smad4	-	+	-			Shg	-	+	-		
ORI-9	Mean Trained					S.E. Trained				
E-cad	0.96	0.00	0.96	1.00		LEF1	0.26	0.00	0.28	0.17	
VIM	0.00	1.00	0.00	0.00		E-cad	0.00	0.11	0.00	0.00	
TGFB2	-	+	-	+	+	Smad	-	-	-	+	
IN Smad4	-	-	+	-		Shg	-	-	+	-	
IN LEF1	-	-	-	+		IN LEF1	-	-	-	+	
ORI-10	Mean Trained					S.E. Trained				
Smad	0.00	1.00	0.00			Shg	0.00	0.19	0.00		
Shg	0.00	1.00	0.00			LEF1	0.00	0.20	0.00		
TGFB2	-	+	+			Smad	-	-	+		
IN MEK	-	-	+			Shg	-	-	+		
ORI-11	Mean Trained					S.E. Trained				
LEF1	0.05	1.00	0.00			LEF1	0.02	0.15	0.00		
TGFB2	-	+	+			Smad	-	+	+		
IN MEK	-	-	+			Shg	-	-	+		

Mean Trained vs. Random= (78% significantly different)

ORI-1	Mean Predicted					S.E. Predicted				
Smad	0.00	1.00	0.00			Smad	0.00	0.16	0.00	
Shg	0.00	0.00	1.00			Shg	0.00	0.00	0.19	
LEF1	0.00	0.66	1.00			LEF1	0.00	0.18	0.32	
Smad	-	-	+			Smad	-	-	+	
Shg	-	+	-			Shg	-	+	-	
ORI-2	Mean Predicted					S.E. Predicted				
E-cad	1.00	0.00	0.00	0.06	0.06	E-cad	0.25	0.00	0.00	0.01	0.02
VIM	0.00	1.00	0.95	0.00	0.00	VIM	0.00	0.15	0.14	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-3	Mean Predicted					S.E. Predicted				
LEF1	0.00	0.77	1.00			LEF1	0.00	0.22	0.28	
Smad	-	-	+			Smad	-	-	+	
Shg	-	+	-			Shg	-	+	-	
ORI-4	Mean Predicted					S.E. Predicted				
LEF1	0.00	0.95	1.00	0.00	0.00	LEF1	0.00	0.10	0.10	0.00	0.00
E-cad	1.00	0.34	0.37	0.00	0.37	E-cad	0.26	0.13	0.14	0.00	0.06
VIM	0.00	0.72	1.00	0.00	0.00	VIM	0.00	0.15	0.15	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-5	Mean Predicted					S.E. Predicted				
B-catenin	0.00	0.00	0.00	0.02	1.00	B-catenin	0.00	0.00	0.00	0.02	0.41
TCF4	1.00	0.04	0.00	0.95	1.00	TCF4	0.01	0.12	0.00	1.00	0.95
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-6	Mean Predicted					S.E. Predicted				
TGFB2	0.00	0.71	1.00	0.31	0.48	TGFB2	0.00	0.12	0.16	0.08	0.11
E-cad	0.52	0.00	0.00	0.92	1.00	E-cad	0.06	0.00	0.00	0.16	0.17
VIM	0.00	1.00	0.94	0.40	0.39	VIM	0.00	0.15	0.15	0.06	0.06
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-7	Mean Predicted					S.E. Predicted				
Smad	0.00	0.00				Smad	0.00	0.00			
Shg	0.00	0.00				Shg	0.00	0.00			
Smad	-	+				Smad	-	+			
ORI-8	Mean Predicted					S.E. Predicted				
LEF1	0.02	1.00	0.00			LEF1	0.01	0.25	0.00		
Smad	-	-	+			Smad	-	-	+		
Shg	-	+	-			Shg	-	+	-		
ORI-9	Mean Predicted					S.E. Predicted				
LEF1	0.86	0.00	0.86	1.00		LEF1	0.19	0.00	0.19	0.22	
E-cad	0.00	1.00	0.00	0.00		E-cad	0.00	0.26	0.00	0.00	
Smad	-	+	-	+		Smad	-	+	-	+	
Shg	-	-	+	-		Shg	-	-	+	-	
IN LEF1	-	-	-	+		IN LEF1	-	-	-	+	
ORI-10	Mean Predicted					S.E. Predicted				
Shg	0.00	1.00	0.00			Shg	0.00	0.18	0.00		
LEF1	0.08	1.00	0.00			LEF1	0.00	0.12	0.00		
Smad	-	+	+			Smad	-	+	+		
Shg	-	-	+			Shg	-	-	+		
ORI-11	Mean Predicted					S.E. Predicted				
LEF1	0.02	1.00	0.00			LEF1	0.01	0.25	0.00		
Smad	-	+	+			Smad	-	+	+		
Shg	-	-	+			Shg	-	-	+		

Cross Validation vs. Random = (59% significantly different)

ORI-1	Mean Random					S.E. Random				
Smad	0.00	1.00	0.00			Smad	0.00	0.00	0.00	
Shg	0.00	0.00	1.00			Shg	0.00	0.00	0.00	
LEF1	0.01	0.87	0.11			LEF1	0.01	0.02	0.03	
Smad	-	-	+			Smad	-	-	+	
Shg	-	+	-			Shg	-	+	-	
ORI-2	Mean Random					S.E. Random				
E-cad	0.97	0.09	0.16	0.17	0.18	E-cad	0.02	0.02	0.03	0.03	0.03
VIM	0.49	0.23	0.16	0.00	0.00	VIM	0.04	0.04	0.03	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-3	Mean Random					S.E. Random				
LEF1	0.40	0.41	0.36			LEF1	0.08	0.04	0.05		
Smad	-	-	+			Smad	-	-	+		
Shg	-	+	-			Shg	-	+	-		
ORI-4	Mean Random					S.E. Random				
LEF1	0.13	0.70	0.17	0.11	0.11	LEF1	0.03	0.04	0.03	0.03	0.03
E-cad	0.98	0.10	0.17	0.15	0.17	E-cad	0.01	0.02	0.03	0.03	0.03
VIM	0.16	0.63	0.14	0.06	0.08	VIM	0.03	0.03	0.03	0.02	0.03
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-5	Mean Random					S.E. Random				
B-catenin	0.00	0.00	0.00	0.00	0.00	B-catenin	0.00	0.00	0.00	0.00	0.00
TCF4	0.00	0.00	0.00	0.00	0.00	TCF4	0.00	0.00	0.00	0.00	0.00
Smad	-	+	-	+	-	Smad	-	+	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-6	Mean Random					S.E. Random				
TGFB2	0.00	0.38	0.15	0.04	0.07	TGFB2	0.00	0.00	0.03	0.02	0.02
E-cad	0.00	0.00	0.00	0.90	0.98	E-cad	0.00	0.00	0.00	0.00	0.00
VIM	0.32	0.76	0.11	0.43	0.20	VIM	0.04	0.04	0.03	0.02	0.02
Smad	-	-	-	+	-	Smad	-	-	-	+	-
Shg	-	-	+	-	+	Shg	-	-	+	-	+
IN LEF1	-	-	-	+	+	IN LEF1	-	-	-	+	+
ORI-7	Mean Random					S.E. Random				
Smad	0.00	0.00				Smad	0.00	0.00			
Shg	0.00	0.00				Shg	0.00	0.03			
Smad	-	+				Smad	-	+			
ORI-8	Mean Random					S.E. Random				
LEF1	0.27	0.61	0.09			LEF1	0.04	0.05	0.02		
Smad	-	-	+			Smad	-	-	+		
Shg	-	+	-			Shg	-	+	-		
ORI-9	Mean Random					S.E. Random				
LEF1	0.00	0.00	0.00	0.00		LEF1	0.00	0.00	0.00	0.00	
E-cad	0.42	0.33	0.09	0.00		E-cad	0.04	0.04	0.03	0.00	
Smad	-	+	-	+		Smad	-	+	-	+	
Shg	-	-	+	-		Shg	-	-	+	-	
IN LEF1	-	-	-	+		IN LEF1	-	-	-	+	
ORI-10	Mean Random					S.E. Random				
Shg	0.00	0.00	0.00			Shg	0.00	0.00	0.00		
LEF1	0.00	0.00	0.00			LEF1	0.00	0.03	0.00		
Smad	-	+	+			Smad	-	+	+		
Shg	-	-	+			Shg	-	-	+		
ORI-11	Mean Random					S.E. Random				
LEF1	0.15	0.86	0.25			LEF1	0.04	0.03	0.04		
Smad	-	+	+			Smad	-	+	+		
Shg	-	-	+			Shg	-	-	+		

Fig. S6: Training and prediction values as a function of condition for the 11 TGF-β objective functions versus a random parameter control.

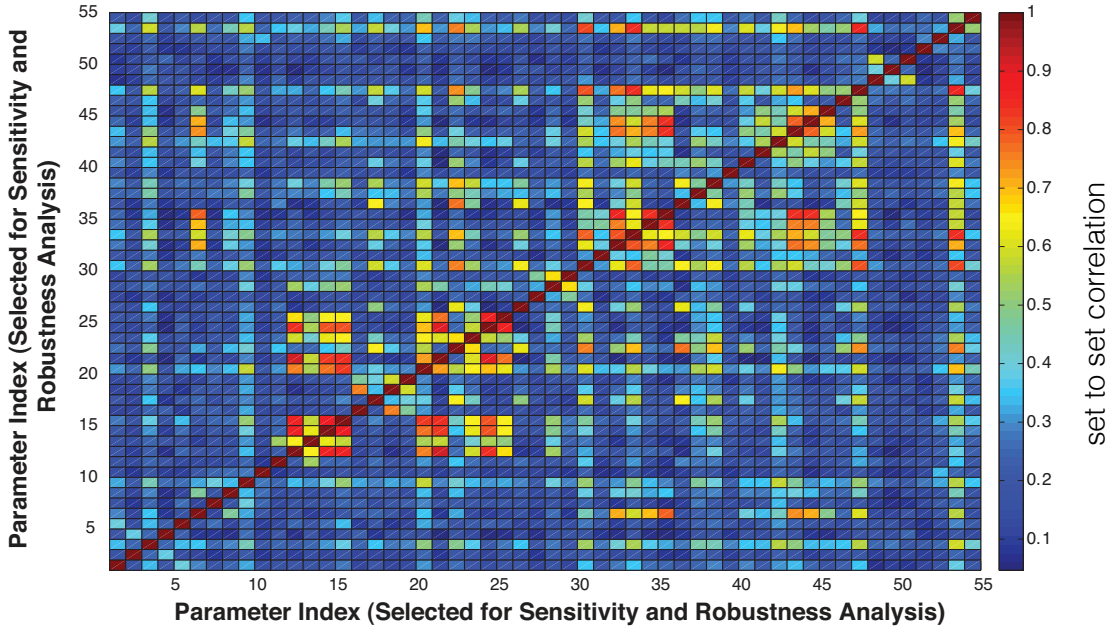

Fig. S7: Parameter set to set correlation for 55 random parameter sets selected from the ensemble. Of the 55 sets selected, the average correlation between sets was less than 25% for greater than 80% of the parameter sets.

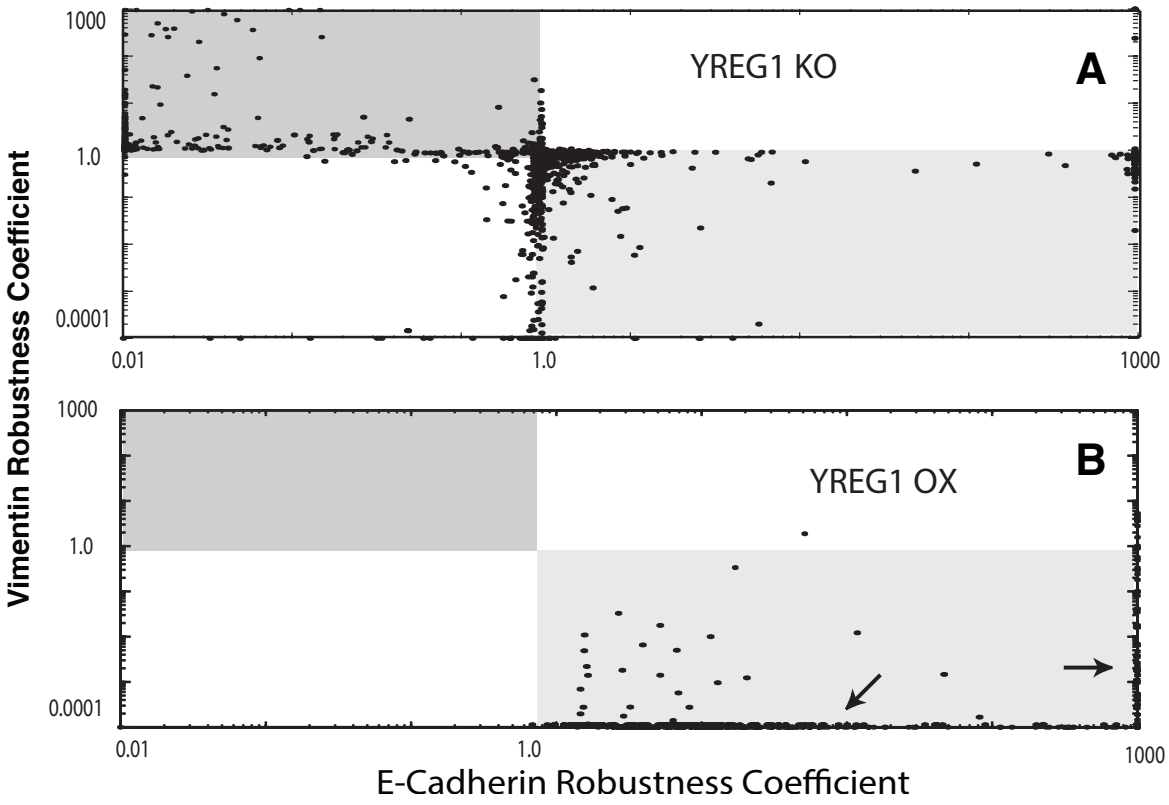

Fig. S8: Robustness of E-cadherin and Vimentin expression to a knockout (A) and overexpression (B) of the hypothetical regulator 1 (YREG1) protein. Robustness coefficients were calculated for each member of the ensemble. Each point represents the response of a single model in the ensemble to either a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

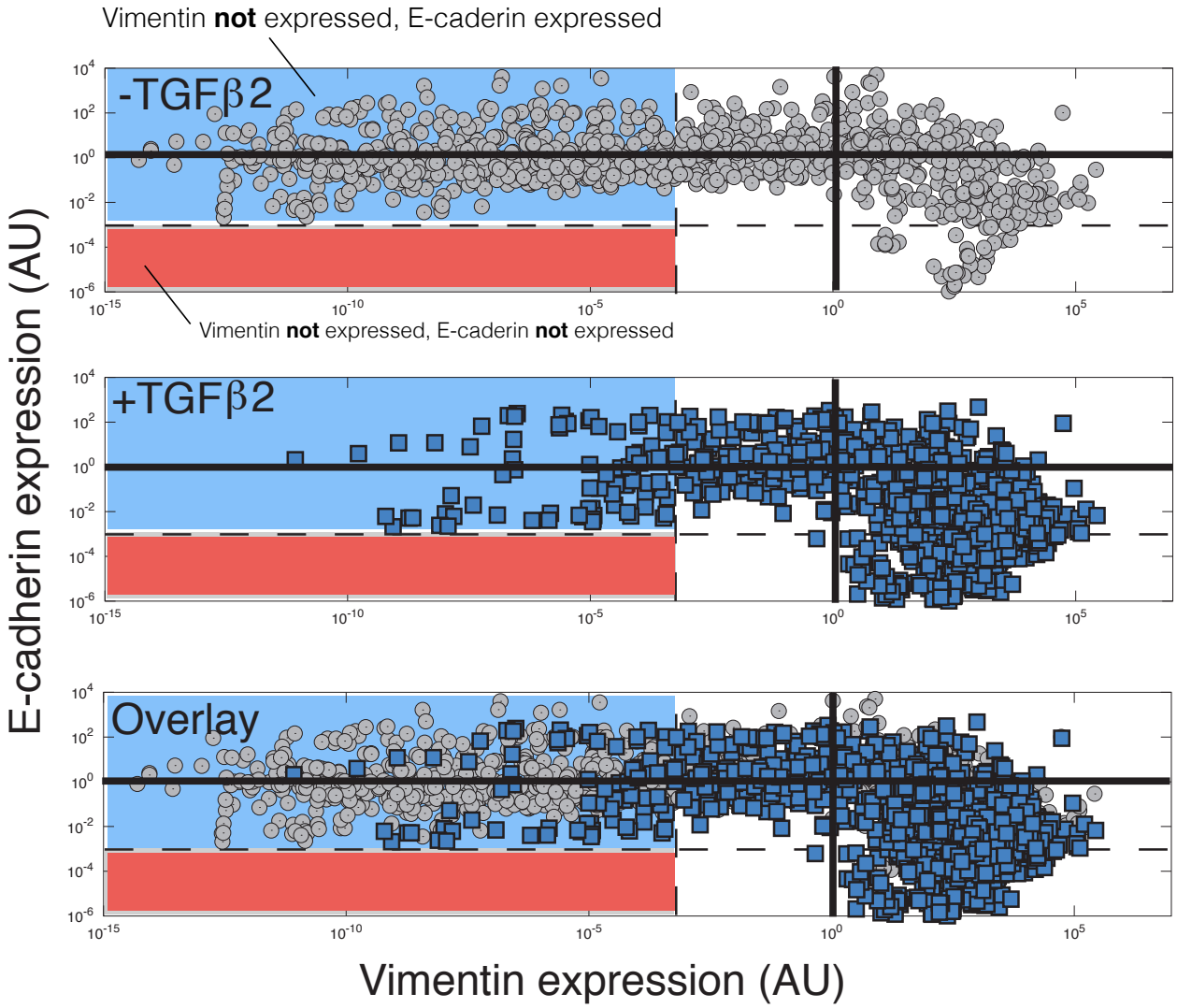

Fig. S9: Steady state protein abundance for E-cadherin and Vimentin (AU) as a function of TGF-β1/2 exposure. Top: Overlay of the model population for Vimentin (AU) and E-cadherin (AU) expression in the presence (blue) and absence (gray) of TGF-β1/2. Midddle: Vimentin (AU) and E-cadherin (AU) expression in the absence of TGF-β1/2 showed exhinibted population heterogenity. Bottom: Vimentin (AU) and E-cadherin (AU) expression in the presence of TGF-β1/2 moved the centriod of the population toward Vimentin (AU) and away from E-cadherin (AU) expression.