

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.

1 Introduction

2 The epithelial to mesenchymal transition (EMT) is a broadly participating, evolutionarily
3 conserved differentiation program essential for tissue morphogenesis, remodeling and
4 pathological processes such as cancer Thiery (2003). During EMT polarized, tightly ad-
5 hered epithelial cell monolayers are transformed into non-interacting motile mesenchymal
6 cells that simultaneously degrade and synthesize extracellular matrix (ECM) components
7 and invade into the underlying tissue space Stahl & Felsen (2001). EMT is the funda-
8 mental initiator of developmental processes such as embryonic gastrulation and valvulo-
9 genesis Eisenberg & Markwald (1995) (also Kalluri J Clin Invest 2009, Thiery Cell 2009).
10 Transforming growth factor β (TGF- β) family members are important inducers of both de-

11 velopmental and pathological EMT Xu *et al.* (2009), Zavadil & Böttinger (2005). Decades
12 of research has focused on identifying molecular regulators of EMT, but almost all on a
13 single gene and in a nearly binary yes/no level of qualitative understanding. Medici and
14 coworkers recently identified a core signaling program by which TGF- β isoforms induce
15 EMT across a variety of cell lines Medici *et al.* (2006, 2008). This program involves care-
16 fully orchestrated rounds of gene expression driven by the Smad and Snail families of
17 transcription factors as well as other key factors such as lymphoid enhancer-binding fac-
18 tor 1 (LEF-1), nuclear factor of activated T-cells, cytoplasmic 1 (NFATc1), and specificity
19 protein 1 (Sp1). Coregulators such as β -catenin, NF- κ B, and the ErbB family of receptor
20 tyrosine kinases however also participate in EMT regulation, but the degree of each's in-
21 fluence is difficult to ascertain in isolation Hardy *et al.* (2010), Huber *et al.* (2004), Jiang
22 *et al.* (2007), Kim *et al.* (2002). EMT also exhibits complex temporal dynamics that are
23 often intractable in gain/loss of function studies. Elucidating the master regulatory ar-
24 chitecture controlling EMT therefore requires inclusion of these complex overlapping and
25 non-binary behaviors.

26 Systems biology and mathematical modeling are essential tools for understanding
27 complex developmental programs like EMT Ahmed & Nawshad (2007). Previous com-
28 putational models of TGF- β induced differentiation focused on single biological factors or
29 EMT in single cells. For example, Chung *et al.*, constructed a model of TGF- β receptor
30 activation and Smad signaling using ordinary differential equations and mass-action ki-
31 netics. Their model suggested that a reduction of functional TGF- β receptors in cancer
32 cells may lead to an attenuated Smad2 signal Chung *et al.* (2009). Similarly, Vilar *et al.*
33 suggested that specific changes in receptor trafficking patterns could lead to phenotypes
34 that favor tumorigenesis Vilar *et al.* (2006). Although these models provided insight into
35 the role of receptor dynamics, EMT induction involves many other components, including
36 competing second messengers and interconnected transcriptional regulatory loops. Inte-
37 grating these additional scales of molecular signaling while maintaining the capacity for

38 robust prediction requires a new and expanded computational and experimental strategy.
39 Data-driven systems approaches Cirit & Haugh (2012) or logical model formulations Mor-
40 ris *et al.* (2011) are emerging paradigms that constrain model complexity through the in-
41 corporation of training and validation data. These are interesting techniques because the
42 data informs model structure (which can be expanded as more data becomes available).
43 Alternatively, Bailey proposed more than a decade ago that a qualitative understanding
44 of a complex biological system should not require complete definition of its structural
45 and parametric content Bailey (2001). Shortly thereafter, Sethna and coworkers showed
46 that complex model behavior is often controlled by only a few parameter combinations, a
47 characteristic seemingly universal to multi-parameter models referred to as “sloppiness”
48 Machta *et al.* (2013). Thus, reasonable model predictions are often possible with only
49 limited parameter information. Taking advantage of this property, we developed sloppy
50 techniques for parameter identification using ensembles of deterministic models Song
51 *et al.* (2010). Furthermore, we proposed that the sloppy behavior of biological networks
52 may also be seen as a source of cell-to-cell Lequieu *et al.* (2011) or even patient-to-patient
53 heterogeneity Luan *et al.* (2010). Recently, Bayesian parameter identification techniques
54 have also been used to explore cell-to-cell heterogeneity Hasenauer *et al.* (2011), Kalita
55 *et al.* (2011), where a population of cells could be viewed as a dynamic ensemble of
56 context-specific biochemical networks Creixell *et al.* (2012).

57 In this study, we developed a family of mechanistic models describing the induction
58 of EMT by TGF- β isoforms in the presence and absence of vascular endothelial growth
59 factor A (VEGF-A). We incorporated mass action kinetics within an ordinary differential
60 equation (ODE) framework to describe the EMT interaction network containing 995 gene,
61 protein or mRNA components interconnected through 1700 interactions. A family of model
62 parameters was estimated using 41 molecular data sets generated in DLD1 colon carci-
63 noma, MDCKII and A375 melanoma cells using the Pareto optimal ensemble technique
64 (POETs) multiobjective optimization algorithm. POETs identified more than 15,000 likely

65 TGF- β induced EMT models, from which we selected approximately 1100 models for fur-
66 ther analysis. Analysis of the model population suggested that both MCF10A and DLD1
67 cells could exhibit phenotypic heterogeneity if treated simultaneously with TGF- β 1/2 and
68 VEGF-A. This heterogeneity was characterized by a significant fraction of the population
69 being in a “hybrid state” having both high E-cadherin and high Vimentin expression. We
70 tested these predictions using qRT-PCR and flow-cytometry studies in a variety of ex-
71 perimental conditions. Validation studies confirmed that upwards of 45% of the cellular
72 population could be put into the hybrid state in the presence of both TGF- β 1/2 and VEGF-
73 A. Moreover, this response depended upon both activation of Sp1 by MAPK and NFATc1
74 transcriptional activity consistent with the predicted molecular signaling. Lastly, the hybrid
75 populations of both DLD1 and MCF10A cells exhibited different functional behavior than
76 those from either TGF- β or VEGF-A treatment. The extent of ductal branch formation
77 significantly increased with MCF10A cells in the hybrid phenotype, compared with cells
78 treated with VEGF-A alone. Together, these results establish a predictive mechanistic
79 model of EMT susceptibility, and reveal a novel signaling axis, which possibly regulates
80 carcinoma progression through an EMT versus tubulogenesis response.

81 Results

82 **The model population captured key features of TGF- β induced EMT** The EMT model
83 architecture, based upon curated molecular connectivity, described the expression of 80
84 genes following exposure to TGF- β isoforms and VEGF-A (Fig. 1). The EMT model
85 contained 995 molecular species interconnected by 1700 interactions. Model equations
86 were formulated using mass-action kinetics within an ordinary differential equation (ODE)
87 framework. ODEs and mass action kinetics are common tools to model biochemical path-
88 ways Chen *et al.* (2009), Schoeberl *et al.* (2002), Tasseff *et al.* (2011). However, while
89 ODE models can simulate complex intracellular behavior, they require estimates for model
90 parameters which are often difficult to obtain. The EMT model had 1756 unknown model
91 parameters, 1700 kinetic constants and 56 non-zero initial conditions. As expected, these
92 parameters were not uniquely identifiable given the training data Gadkar *et al.* (2005).
93 Thus, instead of identifying a single best fit (but uncertain) model, we estimated a sloppy
94 population of models (each consistent with the training data) by simultaneously minimiz-
95 ing the difference between model simulations and 41 molecular data sets using the Pareto
96 Optimal Ensemble Technique (POETs). The training data were generated in DLD1 colon
97 carcinoma, MDCKII, and A375 melanoma cells following exposure to TGF- β isoforms
98 Medici *et al.* (2008). We organized these data sets into 11 objective functions which
99 were simultaneously minimized by POETs. Additionally, we used 12 molecular data sets
100 generated in HK-2 cells following VEGF-A exposure to train VEGF-A responsive model
101 processes Lian *et al.* (2011). To guard against overfitting, we augmented the multiobjec-
102 tive optimization with leave-one-out cross validation to independently estimate both the
103 training and prediction error for each objective. Thus, we generated 11 different model
104 ensembles. Lastly, we compared model predictions with independent data sets not used
105 during training (both at the molecular and model population levels) to evaluate the pre-
106 dictive power of the parameter ensemble. Additional details of the signaling architecture
107 included in the model are presented in the materials and methods and the supplement.

108 POETs generated a population of probable signaling models which captured the mul-
109 tiple phases of EMT induction (Fig. 2). POETs sampled well over 10^6 probable models
110 during each stage of the cross-validation, using a combination of both local and global
111 random sampling. This sampling generated approximately 15,000 highly probable mod-
112 els from which we selected $N \simeq 1100$ models for further analysis. The selected models all
113 had the same possible molecular connectivity, but different values for model parameters
114 and extrinsic factors such as RNA polymerase or ribosome abundance. Model selec-
115 tion was based upon Pareto rank, the prediction and training error across all objectives
116 and model to model correlation (supplemental materials). The model population recapit-
117 ulated key signaling events following TGF- β exposure. We subdivided the response to
118 TGF- β exposure into two phases. First, TGF- β 1/2 signaling initiated a program which
119 downregulated E-cadherin expression in a MAPK dependent manner while simultane-
120 ously upregulating TGF- β 3 expression. Second, TGF- β 3 secretion initiated an autocrine
121 feedback which upregulated the expression of mesenchymal markers such as Vimentin
122 and key upstream transcription factors such as LEF-1 in a SMAD dependent manner.
123 Each phase involved the hierarchal expression and/or post-translational modification of
124 several key transcription factors. During the first phase, stimulation with TGF- β 1/2 (10
125 a.u.) activated both the SMAD and MAPK pathways. MAPK activation resulted in the
126 phosphorylation of the transcription factor activator protein 1 (AP-1), which in-turn upreg-
127 ulated the expression of Snail, a well established transcriptional repressor (Fig. 2A). Snail
128 expression was MAPK-dependent; the MEK inhibitor U0126 blocked AP-1 activation and
129 Snail expression following TGF- β 1/2 exposure (Fig. 2A, Lane 3). Similar results were ob-
130 tained for Slug expression, confirming initial activation through the MAPK pathway (data
131 not shown). Overexpression of either Snail or Slug upregulated TGF- β 3 expression (Fig.
132 2C) while simultaneously downregulating E-cadherin expression (Fig. 2F). During the
133 second phase, TGF- β 3 secretion and the subsequent autocrine signaling resulted in the
134 upregulation of mesenchymal marker expression. The TGF- β 3 induced gene expres-

135 sion program involves a complex hierarchy of transcriptional and post-translational reg-
136 ulatory events. Absence of E-cadherin indirectly promoted TGF- β 3 expression through
137 the β -catenin/TCF4 complex following Snail or Slug expression (Fig. 2C, Lane 2 or 3).
138 Conversely, over-expression of E-cadherin inhibited the TGF- β 3 autocrine production by
139 sequestering cytosolic β -catenin, thereby blocking EMT (Fig. 2C, Lane 4 or 5). TGF- β 3
140 signaled through through the Smad pathway to regulate LEF-1 expression and down-
141 stream target EMT genes (Fig. 2G). TGF- β 3 (10 a.u.) in combination with downstream
142 inhibitors (DN-Smad4 and DN-LEF-1) completely inhibited Vimentin expression, while el-
143 evating E-cadherin expression (Fig. 2H,I).

144 The predictive power of the ensemble was tested using both cross validation and by
145 comparing simulations with data sets not used for model training. In whole, 78% of our
146 training objectives were statistically significant (at a 95% confidence interval) compared to
147 a randomized parameter family (N = 100) generated from the best-fit nominal set (starting
148 point for the optimization). Conversely, we *predicted* approximately 60% of the training
149 objectives, at a 95% confidence interval compared to randomized parameters. The model
150 also captured the temporal gene expression responses of E-cadherin, pSmad2, and LEF-
151 1 to within one-standard deviation (up to the 48 hr time-point) (Fig. 2J-L). This data was
152 not used for model training. The high predictability can be attributed to the combination
153 of the leave-one-out cross validation scheme, diverse objective functions, and robustness
154 of the POETs algorithm. Taken together, the model captured the key signaling events
155 revealed by Medici *et al.* Medici *et al.* (2008) that drive the phenotypic conversion. A
156 listing of data used for training is included in the supplement (Fig. S5 and Fig. S6).

157 **Identification of a novel LEF-1 regulator** During model identification, we found that
158 consistent TGF- β induced EMT required an additional regulatory protein. This protein,
159 which we called hypothetical regulator 1 (YREG1), was required to mediate between
160 SNAIL/SLUG transcriptional activity and the upregulation of LEF-1 expression following
161 TGF- β 1/2 exposure. SNAIL/SLUG are well known transcriptional repressors Dhasarathy

162 *et al.* (2011), Hemavathy *et al.* (2000a,b), although there are a few studies which suggest
163 that at least SNAIL can also act as a transcriptional activator Guaita *et al.* (2002). In
164 the model, we assumed the expression of SNAIL/SLUG was likely regulated by AP1/SP1
165 Jackstadt *et al.* (2013). Thus, upon receiving a TGF- β 1/2 signal, the model predicted en-
166 hanced SNAIL/SLUG expression, consistent with experimental observations. TGF- β 1/2
167 stimulation also induces LEF-1 expression. However, literature evidence suggested that
168 LEF-1 expression was not strongly dependent upon AP1/SP1 activity Eastman & Gross-
169 chedl (1999). Thus, either SNAIL/SLUG are acting as inducers (contrary to substantial
170 biochemical evidence) or, they are repressing the expression of an intermediate repres-
171 sor. Given the biochemical evidence supporting SNAIL/SLUG as repressors, we cre-
172 ated YREG1 a hypothetical intermediate repressor whose expression is downregulated by
173 SNAIL/SLUG. The literature data therefore suggested that YREG1 had two transcriptional
174 targets, LEF-1 and TGF- β 3. By adding this regulator, our simulations became consistent
175 with training and literature data. Medici *et al.* suggested a similar idea where feedback
176 between β -catenin and LEF-1 was likely, although this feedback had yet to be identified
177 Medici *et al.* (2008). Low levels of YREG1 expression were used in all simulations to
178 regulate the formation of the β -catenin-LEF-1 complex. To test the potency of YREG1,
179 we conducted knockdown and over-expression simulations following the addition of TGF-
180 β 1/2 (Fig. S8). In the absence of YREG1, most of the population failed to consistently
181 respond to TGF- β 1/2 exposure compared to the wild-type (Fig. S8A). Conversely, YREG1
182 overexpression revealed an exclusively epithelial phenotype following TGF- β 1/2 stimula-
183 tion (Fig. S8B). Overexpression of YREG1 repressed LEF-1 and TGF- β 3 expression,
184 thereby not allowing free β -catenin to form the β -catenin-LEF-1 complex which promotes
185 mesenchymal gene expression, or SMAD activity following from autocrine TGF- β 3 signal-
186 ing. Likewise, the abundance of the pSmad2/4-LEF-1 complex was also reduced in cells
187 overexpression YREG1, which blocked the repression of E-cadherin. Taken together,
188 we found that low YREG1 expression was necessary for stabilizing EMT, while elevated

189 YREG1 levels limited the extent of EMT induction.

190 **TGF- β 1/2 and VEGF-A exposure promotes phenotype heterogeneity through NFATc**
191 **and phosphorylated Sp1** While we captured the central tendency of many of the molec-
192 ular features of EMT induction following TGF- β 1/2 exposure, an often neglected but im-
193 portant emergent feature of developmental and pathological programs is population het-
194 erogeneity Park *et al.* (2010). We (and others) previously hypothesized that deterministic
195 model ensembles can interrogate population behavior, at least at a course grained level
196 Lequieu *et al.* (2011). We tested this hypothesis by analyzing the response of the pop-
197 ulation of EMT models to extracellular cues and then comparing this response to flow
198 cytometry studies. We used robustness coefficients to quantify the response of the in-
199 dividual members of the ensemble to TGF- β 1/2 stimulation. We have previously used
200 robustness coefficients to systematically quantify response of a system to structural or
201 operational perturbations, for example gene deletions or the addition of a growth factor or
202 hormone Lequieu *et al.* (2011), Song *et al.* (2010), Tasseff *et al.* (2010, 2011). Robust-
203 ness coefficients quantify shifts in molecular marker abundance resulting from molecular
204 of environmental perturbations relative to an unperturbed control state. Robustness coef-
205 ficients $\gg 1$ indicate that marker abundance increased, while robustness coefficients \ll
206 1 indicates marker abundance decreased relative to an unperturbed control. A value of
207 ~ 1 indicates approximately no change in marker abundance following the perturbation.
208 We calculated robustness coefficients for each member of the ensemble ($N \simeq 1100$) for
209 two downstream phenotypic markers, Vimentin (mesenchymal) and E-cadherin (epithe-
210 lial) following the addition of TGF- β 1/2 alone (Fig. 3), and VEGF-A in combination with
211 NFATc inhibitors (Fig. 4). The absence of TGF- β 1/2 or VEGF-A stimulation was used as
212 the baseline for the robustness calculations.

213 We identified model subpopulations that exhibited different behaviors following expo-
214 sure to TGF- β 1/2 (Fig. 3A, labeled P1-P4). Analysis of the molecular signatures of
215 these subpopulations suggested the abundance, localization and state of the Sp1, AP-1
216 and NFATc transcription factors controlled population heterogeneity. The behavior of the

217 majority of models ($>70\%$) was similar to subpopulation one (P1) or subpopulation two
218 (P2) in Fig 3. These models showed the classically expected behavior, a switch from
219 an epithelial to mesenchymal phenotype following TGF- β 1/2 exposure. Models near P1
220 had elevated nuclear localized phosphorylated Sp1, relative to non-induced cells (and
221 models near P2). Elevated Sp1 activity decreased E-cadherin expression through Slug-
222 mediated inhibition, which in turn increased Vimentin expression through TGF- β 3 au-
223 tocrine signaling and the liberation of β -catenin. Near P2, Sp1 transcriptional activity was
224 lower than P1, leading to only modestly increased Vimentin expression and E-cadherin re-
225 pression following TGF- β 1/2 stimulation. Near subpopulation three (P3), reduced levels
226 of nuclear phosphorylated AP-1, Sp1, and NFAT (resulting from the loss of ERK kinase ac-
227 tivity) were responsible for Vimentin *repression* relative to the control. However, the most
228 biologically interesting behavior was exhibited by subpopulation four (P4). Models near
229 P4 had elevated Sp1 and NFAT transcriptional activity, which increased *both* Vimentin and
230 E-cadherin expression. Analysis of these hypothetical cells suggested they had *abnormal*
231 signaling; deregulated NFAT expression and nuclear localization promoted E-cadherin ex-
232 pression while TGF- β 1/2 induced Sp1 action promoted Vimentin expression. Analysis
233 of the connectivity and information flow through the signaling architecture suggested that
234 Sp1 and NFAT action could be manipulated *independently* by simultaneous TGF- β 1/2
235 and VEGF-A stimulation (Fig. S1).

236 To test this hypothesis, we simulated the response of the network to TGF- β 1/2 and
237 VEGF-A treatment with and without NFATc inhibitors (Fig. 4). As expected, stimulation
238 with VEGF-A (50 a.u.) maintained an epithelial population (Q4-43.6%), while TGF- β 1/2
239 (10 a.u.) exposure shifted the population from an epithelial (Q4-5.5%) to a mesenchymal
240 (Q1-45.6%) phenotype (Fig. 4A and Fig. 4B). On the other hand, combined stimula-
241 tion with TGF- β 1/2 (10 a.u.) and VEGF-A (50 a.u.) increased both E-cadherin and
242 Vimentin expression (Q2-45.3%), resulting in a hybrid phenotype with both epithelial and
243 mesenchymal characteristics (Fig. 4C). To better understand this hybrid response, we

244 quantified the simulated protein levels for E-cadherin, Vimentin, phosphorylated nuclear
245 Sp1, nuclear NFATc1, α -smooth muscle actin (α -SMA) and Slug as a function of condition
246 (Fig. S2A-C). Vimentin expression was correlated with high levels of nuclear phosphory-
247 lated Sp1, following TGF- β 1/2 exposure. Conversely, elevated E-cadherin expression
248 depended upon the activity of NFAT transcription factors downstream of VEGF-A stimula-
249 tion. To further isolate the role of NFAT on this hybrid state, we simulated the inhibition of
250 NFAT transcriptional activity across all conditions (all else being equal). NFAT inhibition in
251 combination with VEGF-A treatment blocked all E-cadherin positive sets (Fig. 4D). Like-
252 wise, TGF- β 1/2 treatment in combination with NFATc inhibition also resulted in the loss
253 of E-cadherin expression (Fig. 4E). Lastly, NFATc inhibition in combination with simul-
254 taneous TGF- β 1/2 and VEGF-A exposure repressed nearly all E-cadherin expression,
255 shifting nearly the entire population towards a mesenchymal phenotype (Fig. 4F). Taken
256 together, high levels of nuclear localized phosphorylated Sp1 correlated with Vimentin ex-
257 pression, while NFATc transcriptional activity was predicted to be critical for maintaining
258 E-cadherin expression.

259 **Combined TGF- β 2 and VEGF-A exposure drives heterogeneity in MCF10A and**
260 **DLD1 cells** The EMT model simulations suggested the transcriptional activity of NFATc
261 and Sp1 could be independently tuned to generate a hybrid cell population with both
262 epithelial and mesenchymal characteristics. To test this hypothesis, we exposed either
263 quiescent epithelial (MCF10A, (Fig. 5)) or transformed epithelial cells (DLD1, (Fig. S3))
264 to combinations of TGF- β 1/2 and/or VEGF-A. As expected, treatment with TGF- β 1/2
265 (10ng/ml) increased Slug and Vimentin expression, while repressing E-cadherin expres-
266 sion both at the transcript and protein levels in MCF10A (Fig. 5A-B) and DLD1 cells (Fig.
267 S4C, Fig S3 D,E). Both MCF10A (Fig. 5C) and DLD1 cells (Fig. S3E,G) transitioned
268 from quiescent cobblestone morphology to spread spindle shapes, consistent with EMT.
269 As predicted, we found increased nuclear localization of phosphorylated Sp1 following
270 TGF- β 1/2 stimulation in both MCF10A (Fig. 5B,C) and DLD1 cells (Fig. S3E,F). Con-

271 sistent with model predictions, VEGF-A (50ng/ml) treatment increased the abundance of
272 NFATc1 and E-cadherin at both the transcript and protein level in both MCF10A (Fig.
273 5A) and DLD1 (Fig. S3A) cells. We also found that NFATc1 nuclear localization signif-
274 icantly increased in both MCF10 and DLD1 treated with VEGF-A independently of the
275 abundance of nuclear localized phosphorylated Sp1 levels (Fig. 5B,C Fig.S3C,E). Inter-
276 estingly, combining VEGF-A (50ng/ml) with TGF- β 1/2 (10ng/ml) resulted in significantly
277 elevated expression of both E-cadherin and Vimentin at the transcript and protein levels
278 in both MCF10A and DLD1 cells (Fig 5A,B; Fig S3D,E; Fig S4C). NFATc1 expression
279 increased, while Sp1 expression was similar to the TGF- β 1/2 case alone (Fig. 5A-B,
280 Fig S3D,E; Fig S4C)), supporting their independent regulation. The expression of Slug,
281 and Vimentin significantly increased, while E-cadherin levels were increased in MCF10A
282 cells (Fig 5A) and maintained at control levels in DLD1 cells (Fig. S3D). As further pre-
283 dicted, nuclear co-localization of both NFATc1 and phosphorylated Sp1 were apparent in
284 MCF10A and DLD1 cells treated with both ligands (Fig. 5B,C Fig S3E,F). Taken together,
285 combined VEGF-A and TGF- β 1/2 treatment elicited a hybrid phenotype expressing both
286 mesenchymal and epithelial characteristics in both MCF10A and DLD1 cells. This phe-
287 notype was driven by the transcriptional activity of two key transcription factors, Sp1 and
288 NFATc, which could be modulated independently by TGF- β 1/2 and VEGF-A exposure.

289 Our robustness analysis predicted that NFATc transcriptional activity was critical to
290 maintaining E-cadherin expression in the presence of both VEGF-A and TGF- β 1/2. We
291 experimentally tested this hypothesis by exposing both MCF10A (Fig. 5E,F) and DLD1
292 cells (Fig. S4) to combinations of VEGF-A and TGF- β 1/2 in the presence or absence of
293 VIVIT, a soluble peptide inhibitor of NFATc transcriptional activity Aramburu *et al.* (1999).
294 Treatment with VEGF-A (50ng/ml) and VIVIT (10 μ M) in MCF10A cells significantly re-
295 duced E-cadherin expression compared to VEGF-A alone (Fig 5D,E). Co-treatment with
296 VIVIT and TGF- β 1/2 did not enhance EMT capacity of MCF10A cells above that of
297 TGF- β 1/2 alone (Fig 5A,B,E). Likewise, VIVIT in combination with both TGF- β 1/2 and

298 VEGF-A resulted in a loss of E-cadherin gene and protein expression, while Slug and
299 Vimentin levels remained increased (Fig. 5D,E). Quantitative flow cytometry confirmed
300 these results in both MCF10A (Fig. 5F) and DLD1 cells (Fig. S4C). Both epithelial cell
301 lines initially had high levels of E-cadherin expression, and low vimentin abundance (Q1-
302 99.5%), but both MCF10A and DLD1 cells shifted from an epithelial to mesenchymal
303 phenotype (Q1-33.4%, Q4-42.8%) following TGF- β 1/2 exposure. As expected, NFATc
304 nuclear localization was repressed with VIVIT treatment regardless of ligand stimulation,
305 while the abundance of nuclear phosphorylated Sp1 increased for both TGF- β 1/2 and
306 TGF- β 1/2 + VIVIT conditions (Fig. 5D,E). Combined TGF- β 1/2 and VEGF-A increased
307 both Vimentin and E-cadherin expression (Q1-42.1%, Q2-52.3%) compared to TGF- β 1/2
308 alone. Together, these results demonstrate that NFATc and phosphorylated Sp1 are critical
309 for regulating E-cadherin and Vimentin expression during phenotype heterogeneity in
310 MCF10A and DLD1.

311 **Ductal branching during acini formation is dependent upon phenotype heterogeneity**
312 **in MCF10A and DLD1 cells** We finally employed established three-dimensional
313 (3D) *in vitro* models of invasion, migration, compaction, and tubulogenesis Dhimolea
314 *et al.* (2010) to determine the functional consequences of the hybrid phenotype (Fig. 6).
315 MCF10A and DLD1 cells were aggregated via hanging drop, placed on the surface of a
316 collagen gel, and cultured for 72 hrs under various biochemical treatments. TGF- β 1/2
317 stimulation significantly enhanced cell matrix invasion and matrix compaction, while in
318 contrast VEGF-A stimulation promoted surface migration but no invasion or compaction
319 (Fig. 6B-D). Interestingly, combined TGF- β 1/2 and VEGF-A stimulation significantly in-
320 creased cell migration potential above that of VEGF-A alone while maintaining 3D matrix
321 compaction, though with decreased magnitude compared to TGF- β 1/2 alone. Inhibi-
322 tion of NFATc transcriptional activity by VIVIT decreased migration following treatment
323 with VEGF-A alone (Fig. 6B). Co-treatment of VIVIT significantly decreased migration,
324 while complementarily increasing invasion and compaction, when MCF10A cells were

325 stimulated with both VEGF-A and TGF- β 1/2 (Fig. 6B-D). The responses of DLD1 cells
326 followed a similar trend to MCF10A, although the magnitudes of migration, invasion, and
327 compaction were less. Cell circularity within 3D gels strongly and negatively correlated
328 with both invasion and compaction regardless of treatment (Fig. 6E). Circularity refers
329 to the morphology of the cells. In general, a quiescent epithelial cells assumes a circular
330 morphology in culture, while an active mesenchymal cell is highly elongated. The circular-
331 ity index, a common means of quantifying cell morphology, relates cell area to perimeter.
332 A perfect circle has a circularity index equal to 1.0, while a straight line has a circularity
333 index equal to 0.0, see Butcher et al. Butcher *et al.* (2004). TGF- β 1/2 treatment alone
334 resulted in irregular and spindle shaped morphology, while VEGF-A exposure promoted
335 round quiescent cells (Fig. 6A). Combined VEGF-A and TGF- β 1/2 promoted morphology
336 between these extremes. VIVIT mediated NFATc inhibition significantly reduced the cir-
337 cularity index, similar to TGF- β 1/2 treatment (Fig. 6F). VEGF-A treatment also induced
338 the formation of tubular structures (acini), but the number of tubular branches relative to
339 total acini was significantly increased upon combined TGF- β 1/2 and VEGF-A. No tubular
340 structures were identified within the DLD1 constructs during the 7 day tubulogenesis end-
341 points, supporting that MCF10A and DLD1 cells have some cell-type specific EMT sensi-
342 tivity despite their underlying competency for acquiring a heterogeneous phenotype. This
343 suggests that initial EMT sensitivity of a cell influences downstream functional response
344 from TGF- β and VEGFA stimulation. Together, these results establish that VEGF-A and
345 TGF- β 1/2 ligand concentrations potentiate between acini and ductal branch formation in
346 3D culture, and are dependent upon NFATc activity.

347 Discussion

348 In this study, we developed a family of mechanistic models describing the induction of
349 EMT by TGF- β isoforms in the presence and absence of VEGF-A. The signaling archi-
350 tecture encoded in the model, which contained 995 molecular species interconnected
351 by 1700 interactions, described the expression of 80 genes in response to growth fac-
352 tor stimulation. This simulation incorporates an unprecedented level of detail compared
353 to previous models, but as a consequence created a large number of unknown model
354 parameters. Because these parameters could not be estimated uniquely a priori, we es-
355 timated an ensemble of likely parameters using the POETs multiobjective optimization
356 framework. The model population was trained and cross-validated to prescribe biolog-
357 ical significance using 41 data sets generated in DLD1 colon carcinoma, MDCKII, and
358 A375 melanoma cell lines Medici *et al.* (2008). POETs generated $> 15,000$ probable pa-
359 rameter sets using this data, from which we selected $N \simeq 1100$ for subsequent analysis.
360 Analysis of this population predicted possible phenotypic modes (and their associated
361 signaling) that cells could exhibit when stimulated with TGF- β and/or VEGF-A. The most
362 novel hypothesis generated from the analysis was that cells could operate in a hybrid
363 state defined by both epithelial and mesenchymal traits when stimulated simultaneously
364 with TGF- β and VEGF-A. We tested this hypothesis in MCF10A and DLD1 cells stimu-
365 lated with combinations of TGF- β and VEGF-A. As expected, in the presence of TGF- β
366 or VEGF-A alone, MCF10A and DLD1 cells were either mesenchymal or epithelial, re-
367 spectively. However, with both TGF- β and VEGF-A, MCF10A and DLD1 cells exhibited a
368 hybrid phenotype, having both epithelial and mesenchymal characteristics. Furthermore,
369 we found that functional traits such as tubulogenesis and ductal branching were different
370 for cells in this hybrid phenotype. Together, this study established a predictive model of
371 EMT induction, determined that deterministic model ensembles could predict population
372 heterogeneity, and proved the existence of a unique hybrid phenotype resulting from the
373 simultaneous integration of extracellular growth factor signals.

374 Cells routinely process a multitude of signals simultaneously, especially when coordi-
375 nating developmental or pathological programs. For example, oncogenic cells integrate
376 both mechanical and chemical cues in their local microenvironment during tumorigenesis,
377 including cytokines VEGF and TGF- β Hong *et al.* (2013). VEGF-A mediates patholog-
378 ical angiogenic remodeling of tumors Nagy *et al.* (2007), while TGF- β can elicit both
379 protective and oncogenic responses Ferrara (2002), Willis & Borok (2007). While much
380 research has tested signaling pathways individually, far less is understood about combi-
381 natorial stimulation, such as with both VEGF-A and TGF- β . Recent *in vitro* and *in vivo*
382 evidence has suggested that epithelial cells can exhibit heterogeneous phenotypes in ad-
383 dition to classically defined epithelial or mesenchymal states Polyak & Weinberg (2009),
384 Strauss *et al.* (2011). For example, expression profiling in human epithelial cancer cell
385 lines demonstrated a spectrum of phenotypes, including some that expressed both E-
386 cadherin and Vimentin simultaneously Neve *et al.* (2006), Welch-Reardon *et al.* (2014).
387 Zajchowski *et al.*, speculated that these expression profiles were somehow important for
388 maintaining epithelial properties, while simultaneously allowing other functional behavior
389 such as proliferation and migration Zajchowski *et al.* (2001). Whether and how heteroge-
390 neous phenotypes arise and participate in cancer progression, as well as their response
391 to pharmacological inhibition are fundamental questions that should receive increased at-
392 tention. In this study, we determined that a hybrid phenotype could be obtained through
393 combined treatment with VEGF-A and TGF- β , both common factors localized in the tu-
394 mor microenvironment. Furthermore, our systematic simulation-experimentation strategy
395 identified that the transcriptional activity of Sp1 and NFATc were the critical factors con-
396 trolling this phenotypic heterogeneity. Several studies have highlighted the importance
397 of NFATc as a key transcription factor involved in cell growth, survival, invasion, angio-
398 genesis and cancer Mancini & Toker (2009). For example, proliferation and anchorage-
399 independent growth of pancreatic tumor cells is dependent on calcineurin and NFATc1
400 activity, consistent with the high levels of nuclear NFATc1 found in pancreatic tumors

401 Singh *et al.* (2010). Likewise, our results found that VEGF-A was a potent inducer of
402 NFATc1 expression, which may be required for epithelial cell migration and tubulogene-
403 sis. Although specific NFATc isoforms were not distinguished in the model, our simulations
404 suggested that NFATc transcriptional activity was capable of maintaining epithelial traits,
405 even during TGF- β induced EMT. Experimentally, we found that E-cadherin expression
406 was dependent upon NFATc dephosphorylation in response to simultaneous VEGF-A and
407 TGF- β 1/2 treatment. Thus, these results support the hypothesis that NFATc activity plays
408 a critical role in maintaining cell-cell contacts, even during partial EMT.

409 Epithelial cells reproduce tissue-like organization when grown in a three-dimensional
410 extracellular matrix (ECM) environment, and therefore are an attractive model to study
411 morphogenic mechanisms. It is well established that MCF10A cells form structures that
412 closely resemble acini (multi-lobed cluster of cells) in three-dimensional *in vitro* cultures
413 Debnath *et al.* (2003). It has been postulated that a cellular response reminiscent of partial
414 EMT underlies this process, stimulating further branching and formation of acini Pearson
415 & Hunter (2007). Normally well controlled process such as tubulogenesis can be co-opted
416 by cancer cells to break away from a primary lesion and invade through the surrounding
417 stroma O'Brien *et al.* (2004). However, by retaining a transient hybrid EMT-like state, clus-
418 ters of these tube-forming tumor cells can reform at a high rate after invasion, possibly
419 explaining why invasive human carcinomas frequently appear to be cellular collections
420 with varying degrees of gland-like differentiation Debnath & Brugge (2005). In this study,
421 we showed that our predicted hybrid phenotype generated by simultaneous treatment
422 of epithelial cells with VEGF-A and TGF- β possessed altered migration and invasion,
423 which enhanced tubular branching. A salient feature of this behavior, however, was the
424 retention of cell-cell contacts that allowed cells to migrate without completely dissociat-
425 ing from their neighbors. Thus, our results support a mechanism in which hybrid cells
426 can maintain some functional characteristics of epithelial cells such as cell-cell adhesion,
427 which are normally lost in a fully differentiated mesenchymal state. The tumor microen-

428 vironment contains many soluble signals simultaneously, including VEGF and TGF- β .
429 Thus, it is likely that some cancerous epithelial cells could exhibit hybrid EMT pheno-
430 typic states. This may explain why fibroblastoid morphology, a classical feature of EMT,
431 is not commonly observed in human carcinomas Debnath & Brugge (2005). This study
432 focused on the combinatorial effects of two very different ligand families present together
433 in the tumor environment. Additional modeling studies are required to unravel the global
434 response of epithelial cells to the full spectrum of chemical, substrate, and mechanical
435 cues. The simulation strategy presented here is readily adaptable to larger species sets,
436 with the major advantage that experimentally testable hypotheses can be generated re-
437 garding how signals get integrated to produce global cellular response. Furthermore, by
438 simulating multiple ensembles of parameter sets, subpopulations across a constellation of
439 phenotypes can be created and mined for common and/or divergent signaling character-
440 istics. This is a significant advantage over forced convergence to a single unique solution
441 and thereby generating a potentially non-physiological homogeneous population.

442 The deterministic population of EMT models predicted heterogeneous behavior that
443 was qualitatively consistent with experimental studies. There is a diversity of algorithmic
444 approaches to estimate model parameters Moles *et al.* (2003), as well as many strate-
445 gies to integrate model identification with experimental design Rodriguez-Fernandez *et al.*
446 (2013), Villaverde & Banga (2014). However, despite these advances, the identification
447 of models describing intracellular network behavior remains challenging. There are dif-
448 ferent schools of thought to deal with this challenge. One school has focused on model
449 reduction. Data-driven approaches Cirit & Haugh (2012), boolean Choi *et al.* (2012) or
450 other logical model formulations Morris *et al.* (2011), Terfve *et al.* (2012) are emerging
451 paradigms that constrain model complexity by the availability of the training and validation
452 data. Other techniques such as constraints based modeling, which is commonly used
453 to model metabolic networks, have also been applied to model transcriptional networks,
454 although primarily in lower eukaryotes and prokaryotes Hyduke & Palsson (2010). These

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

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

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

509 in epithelial transformation. NF- κ B may influence Snail expression through the AKT path-
510 way and directly stabilize Snail activity Wu *et al.* (2009). This is particularly important
511 for integrating inflammation pathways, such as interleukin-6 (IL-6) and tumor necrosis
512 factor- α (TNF- α), which have been linked to EMT in pathological conditions Sullivan *et al.*
513 (2009). Other pathways such as Notch have also been shown to act synergistically with
514 TGF- β to express Slug in the developing embryo Niessen *et al.* (2008). Lastly, while we
515 have modeled classical protein signaling, we have not considered the role of regulatory
516 RNAs on EMT. There is growing evidence that microRNAs (miRNAs) play a strong role
517 in EMT, where several miRNAs, for example miR-21 and miR-31 are strongly associated
518 with TGF- β exposure Bullock *et al.* (2012). Addressing missing structural components
519 like these, could generate more insight into TGF- β signaling and its role in phenotypic
520 transformation.

521 **Materials and Methods**

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

524 **Signaling network connectivity** The EMT model described the gene expression pro-
525 gram resulting from TGF- β and VEGF-A signaling in a prototypical epithelial cell. The
526 TGF- β -EMT network contained 995 nodes (proteins, mRNA or genes) interconnected by
527 1700 interactions. The network connectivity was curated from more than 40 primary liter-
528 ature sources in combination with on-line databases Jensen *et al.* (2009), Linding *et al.*
529 (2007). The model interactome was not specific to a single epithelial cell line. Rather, we
530 assembled canonical pathways involved in TGF- β and VEGF-A signaling, defaulting to
531 human connectivity when possible. Using a canonical architecture allowed us to explore
532 general features of TGF- β induced EMT without cell line specific artifacts. On the other
533 hand, because of the canonical architecture, we evaluated the simulation conclusions in
534 several cell lines to test the generality of our conclusions.

535 Our signaling network reconstruction was based on Medici *et al.* who identified the
536 pathways through which MDCKII, DLD1 colon carcinoma, and A375 melanoma cells tran-
537 sition towards a mesenchymal phenotype Medici *et al.* (2008). Sequential activation of
538 MAPK and Smad pathways were initiated upon addition of TGF β 1/2. Briefly, TGF β 2 sig-
539 nals through the RAS-RAF-MEK-ERK pathway to up-regulate Snail and Slug expression
540 Medici *et al.* (2006). Snail, a known repressor of junctional proteins, inhibits the expres-
541 sion of E-cadherin Cano *et al.* (2000). This initial repression of E-cadherin leads to a
542 release of β -catenin from the cell membrane. Cytosolic β -catenin can then translocate to
543 the nucleus and form transcriptional complexes with TCF-4 to drive TGF β 3 expression
544 Medici *et al.* (2008). TGF β 3 signals to the cells interior by binding to type II receptors,
545 which form heterodimers with type I receptors (ALK5) Derynck & Zhang (2003). This
546 activates the receptors serine/threonine kinase activity to phosphorylate and activate the
547 receptor Smads 2/3 Massagué *et al.* (2005). Phosphorylated Smads 2/3 (pSmad2/3)

548 form heterodimers with partner Smad4 and translocate to the nucleus. pSmads com-
549 plexes up-regulate other transcription factors, such as LEF-1. The pSmad2/4/LEF-1 has
550 been shown to directly repress the E-cadherin gene Nawshad *et al.* (2007). LEF-1 also
551 binds with β -catenin to upregulate mesenchymal proteins such as fibronectin Medici *et al.*
552 (2011). The EMT gene expression program was initiated by the binding of TGF- β iso-
553 forms to TGF- β surface receptors. Binding of extracellular TGF- β 1/2 with TGF- β sur-
554 face receptors I/II (TGF- β R-I/II) initiates the assembly of adapter complexes which starts
555 the downstream signaling program. In the model, TGF- β 1/2 binds TGF- β R-I/II followed
556 by the recruitment of activin receptor-like kinase 1 (ALK1) and TGF- β surface receptor
557 III (TGF- β R-III) to form the activated receptor complex Derynck & Zhang (2003). Alter-
558 natively, we also included activin receptor-like kinase 5 (ALK5) recruitment in combination
559 with Endoglin and TGF- β R-III as a second (redundant) activated receptor complex Gatzka
560 *et al.* (2010). Complex assembly activates the serine/threonine kinase activity on the re-
561 ceptor, leading to the recruitment and phosphorylation of Smad partners Massagué *et al.*
562 (2005). Phosphorylated Smads2/3 (pSmad2/3) form heterodimers with partner Smad4
563 and then translocate to the nucleus where they act as both transcriptional activators and
564 repressors. Nuclear pSmad2/3-Smad4 form transcriptional complexes with several genes
565 in the model including lymphoid enhancer-binding factor 1 (*LEF-1*), Nuclear factor of ac-
566 tivated T-cells, cytoplasmic 1 (*NFACT1*), and Specificity Protein 1 (*SP1*). On the other
567 hand, nuclear pSmad2/3-Smad4 represses (in combination with the LEF-1 protein) the
568 expression of E-cadherin (*Cdh1*) Nawshad *et al.* (2007) and Cadherin 5, type 2 (VE-
569 Cadherin encoded by *Cdh5*). Repression of E-cadherin expression is the central event in
570 the transition from an epithelial to a mesenchymal phenotype Cano *et al.* (2000). How-
571 ever, this transition is not solely driven by transcriptional events. At the protein level, the
572 repression of E-cadherin leads to a release of β -catenin from cell membrane. Cytosolic β -
573 catenin then translocates to the nucleus and forms transcriptionally-active complexes with
574 immunoglobulin transcription factor 2 (TCF-4) to drive TGF- β 3 expression Medici *et al.*

575 (2008). Simultaneously, ERK1/2-mediated phosphorylation of the AP1 and Sp1 transcrip-
576 tion factors can also regulate transcriptional complexes involving NFAT, Slug, and Smads.
577 Lastly, canonical pathways for processing extracellular VEGF-A, BMP and Wnt signals,
578 in addition to the PI3K pathway were also included in the model. Additional information
579 about the interactions included in the model, along with the Systems Biology Markup Lan-
580 guage (SBML) file encoding these interactions are included in the supplemental materials.

581 **Formulation, solution and analysis of the EMT model equations** EMT was modeled
582 using mass-action kinetics within an ordinary differential equation (ODE) framework:

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

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

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

594 The set $\{\mathbf{R}_q\}$ denotes reactants for reaction q , while σ_{jq} denotes the stoichiometric coeffi-
595 cient (element of the matrix \mathbf{S}) governing species j in reaction q . The quantity k_q denotes
596 the rate constant (unknown) governing reaction q . Model equations were generated in the

597 C-programming language using the UNIVERSAL code generator, starting from an text-
598 based input file (supplemental materials). UNIVERSAL, an open source Objective-C/Java
599 code generator, is available as a Google Code project ([http://code.google.com/p/universal-
600 code-generator/](http://code.google.com/p/universal-code-generator/)). Model equations were solved using the CVODE solver in the SUNDI-
601 ALS library Hindmarsh *et al.* (2005) on an Apple workstation (Apple, Cupertino, CA) as
602 previously described Tasseff *et al.* (2011).

603 *Estimation of model parameters using multiobjective optimization.* The EMT model had
604 1756 unknown parameters (1700 kinetic constants and 56 non-zero initial conditions)
605 which were not uniquely identifiable given the training data. Instead, we estimated a pop-
606 ulation of likely models (each consistent with the training data) using 41 data sets gen-
607 erated in DLD1 colon carcinoma, MDCKII, and A375 melanoma cells taken from Medici
608 *et al.* Medici *et al.* (2008). We used the Pareto Optimal Ensemble Technique (POETs)
609 multiobjective optimization framework in combination with leave-one-out cross-validation
610 to estimate an ensemble of model parameters Song *et al.* (2010). Cross-validation was
611 used to calculate both training and prediction error during the parameter estimation pro-
612 cedure Kohavi (1995). The 41 intracellular protein and mRNA data-sets used for iden-
613 tification were organized into 11 objective functions. These 11 objective functions were
614 then partitioned, where each partition contained ten training objectives and one validation
615 objective. The training and validation data were Western blots. Thus, all model simula-
616 tions were in arbitrary units. However, POETs does allow a soft constraint on the order
617 of magnitude of the model concentration scale. In this study, we assumed the natural
618 model concentration scale was pmol/L. We did not place a lower bound on model states.
619 However, based on the pmol/L natural scale, we treated all values less than 10^{-3} as zero
620 (or no expression).

621 **Robustness coefficients.** Robustness coefficients were calculated as shown previously
622 Lequieu *et al.* (2011), Tasseff *et al.* (2011). Robustness coefficients denoted by $\alpha(i, j, t_o, t_f)$:

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

623 quantify the response of a marker to a structural or operational perturbation to the net-
624 work architecture. Here t_o and t_f denote the initial and final simulation time respectively,
625 while i and j denote the indices for the marker and the perturbation respectively. A value
626 of $\alpha(i, j, t_o, t_f) > 1$, indicates increased marker abundance, while $\alpha(i, j, t_o, t_f) < 1$ indi-
627 cates decreased marker abundance following perturbation j . If $\alpha(i, j, t_o, t_f) \sim 1$ the j th
628 perturbation does not influence the abundance of marker i . Robustness coefficients were
629 calculated for each member of the ensemble ($N \simeq 1100$).

630 **Cell culture and experimental interrogation** DLD1 colon carcinoma, MCF10A, and
631 HUVEC were acquired from the American Tissue Culture Collection (Manassas, VA).
632 Cells were grown in culture with RPMI 1640 medium with 10% fetal bovine serum and
633 1% penicillin/streptomycin for DLD1, EBM-2 supplemented with EGM-2, 5% fetal bovine
634 serum, and 1% penicillin/streptomycin for HUVEC, or MGEM 2 supplemented with insulin,
635 bovine pituitary extract, cholera toxin, hEGF, hydrocortisone, 5% horse serum, and 1%
636 penicillin/streptomycin for MCF10A. Cells were serum starved for 24 hours and removed
637 from all experimental conditions. Recombinant VEGFA165 was also removed from cul-
638 ture medium prior to experimentation. Recombinant human TGF- β 2 (R & D Systems,
639 Minneapolis, MN) was added to the culture medium at a concentration of 10 ng/ml and re-
640 combinant VEGFA165 at a concentration of (5ng/ml, 50ng/ml) for all relative experiments.
641 NFAT inhibitor (VIVIT peptide) (EMDBiosciences, Darmstadt, Germany), was added to
642 the culture medium at a concentration of 10 μ M for all relative experiments. Cells were
643 passaged 1:3 or 1:4 every 3-6 d and used between passages 4 and 8.

644 *RT-PCR* RNA extractions were performed using a Qiagen total RNA purification kit (Qi-
645 agen, Valencia, CA) and RNA was reverse transcribed to cDNA using the SuperScript
646 III RT-PCR kit with oligo(dT) primer (Invitrogen). Sufficient quality RNA was determined
647 by an absorbance ratio A260/A280 of 1.8-2.1, while the quantity of RNA was determined
648 by measuring the absorbance at 260nm (A260). Real-time PCR experiments were con-
649 ducted using the SYBR Green PCR system (Biorad, Hercules, CA) on a Biorad CFX96
650 cycler, with 40 cycles per sample. Cycling temperatures were as follows: denaturing,
651 95C; annealing, 60C; and extension, 70C. Primers were designed to detect GAPDH, E-
652 cadherin, vimentin, Slug, Sp1, and NFATc1 in cDNA clones: Sp1 (F-TTG AAA AAG GAG
653 TTG GTG GC, R-TGC TGG TTC TGT AAG TTG GG, Accession NG030361.1), NFATc1
654 (F-GCA TCA CAG GGA AGA CCG TGT C, R-GAA GTT CAA TGT CGG AGT TTC TGA
655 G, Accession NG029226.1). GAPDH, E-cadherin, vimentin, and Slug primers were taken
656 from previously published literature Medici *et al.* (2008).

657 *Antibody Staining* Samples were fixed in 4% PFA overnight at 4C. Samples were then
658 washed for 15 minutes on a rocker 3 times with PBS, permeabilized with 0.2% Triton-X
659 100 (VWR International, Radnor, PA) for 10 minutes, and washed another 3 times with
660 PBS. Samples were incubated overnight at 4C in a 1% BSA (Rockland Immunochemi-
661 cals, Inc., Gilbertsville, PA) blocking solution followed by another 4C overnight incubation
662 with either rabbit anti-human E-cadherin 1:100 (Abcam, ab53033), mouse anti-human
663 phospho-Sp1 1:100 (Abcam, ab37707), mouse anti-human vimentin 1:100 (Invitrogen,
664 V9), and rabbit anti-human NFATc1 (Santa Cruz, sc-7294) 1:100. After 3 washes for 15
665 minutes with PBS, samples were exposed to Alexa Fluor 488 or 568 conjugated (Invit-
666 rogen), species specific secondary antibodies at 1:100 in 1% BSA for 2 hours at room
667 temperature. Three more washes with PBS for 15 minutes were followed by incubation
668 with either DRAQ5 far red nuclear stain (Enzo Life Sciences, Plymouth Meeting, PA) at
669 1:1000.

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

680 *Three-Dimensional Culture and Tubulogenesis Assays* For invasion/migration assays,
681 cells were resuspended in culture media, and allowed to aggregate overnight in hanging
682 drop culture (20 μ L; 20,000 cells). The spherical aggregates were placed on the surface of
683 neutralized type I collagen hydrogels (1.5mg/mL) and allowed to adhere for 2 hrs before
684 adding treatments. Cultures were maintained for 72 hrs, after which they were fixed in
685 4% PFA and slowly rehydrated using PBS. For compaction assays, cells were pelleted
686 via centrifugation and resuspended within a neutralized collagen hydrogel (1.5mg/mL)
687 solution at a density of 400,000 cells/mL. 250 μ L of gel was inoculated into culture wells,
688 which solidified after 60min. Treatments were then added within 800 μ L of the culture
689 medium without serum. Gels were liberated from the surfaces of the culture wells the
690 next day and cultured free floating for an additional 3-7 days, exchanging serum free
691 media with appropriate factors every 48 hrs.

692 Tubulogenesis was defined as a typical nonmalignant acini structure. This includes
693 a polarized epithelial cell, hollow lumen, and the basal sides of the cell are surrounded
694 by ECM proteins (Fig. 6A, Controls or VEGF treated). Previous work has shown that
695 change in the morphological characteristics of nontumorigenic MCF10A epithelial acini
696 occur over time and exploiting them to growth in 3D culture can be quantified. For exam-

697 ple, using image segmentation, Chang et al. Chang *et al.* (2007) examined the elongation
698 of the MCF10A acini at 6, 12, and 96 hours after a particular treatment. Polizzotti et al.
699 Polizzotti *et al.* (2012) also suggested a computational method to quantify acini structure
700 based on morphological characteristics in nonmalignant, noninvasive, and invasive con-
701 ditions. Adapted from these approaches, we first fluorescently labeled our cultures and
702 captured the acini structures by 3D confocal microscopy. Next individual acini structures
703 in the images were segmented by imageJ and labeled. We then extracted the number of
704 ductal branches. Ductal branching was defined as any elongated cell cluster extending
705 away from the total acini structure, which was manually segmented and counted using Im-
706 ageJ. A total of 5 images for each condition were used, and approximately 12 acini were
707 analyzed in each image. Total branching was normalized to the amount of acini present,
708 and provides an overall general assessment to the extent of acini remodeling.

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

715 **References**

- 716 Abramoff M, Magelhaes P, Ram S (2004) Image Processing with ImageJ. *Biophotonics*
717 *International*, **11**: 36–42
- 718 Ahmed S, Nawshad A (2007) Complexity in interpretation of embryonic epithelial-
719 mesenchymal transition in response to transforming growth factor-beta signaling. *Cells*
720 *Tissues Organs* **185**: 131–45
- 721 Aramburu J, Yaffe MB, López-Rodríguez C, Cantley LC, Hogan PG, Rao A (1999) Affinity-
722 driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. *Science*
723 **285**: 2129–33
- 724 Arce L, Pate KT, Waterman ML (2009) Groucho binds two conserved regions of LEF-1 for
725 HDAC-dependent repression. *BMC Cancer* **9**: 159
- 726 Bailey JE (2001) Complex biology with no parameters. *Nat Biotechnol* **19**: 503–4
- 727 Bonni S, Wang HR, Causing CG, Kavsak P, Stroschein SL, Luo K, Wrana JL (2001) TGF-
728 beta induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN
729 for degradation. *Nat Cell Biol* **3**: 587–595
- 730 Bullock MD, Sayan AE, Packham GK, Mirnezami AH (2012) MicroRNAs: critical regula-
731 tors of epithelial to mesenchymal (EMT) and mesenchymal to epithelial transition (MET)
732 in cancer progression. *Biol Cell* **104**: 3–12
- 733 Butcher JT, Penrod AM, García AJ, Nerem RM (2004) Unique morphology and focal ad-
734 hesion development of valvular endothelial cells in static and fluid flow environments.
735 *Arterioscler Thromb Vasc Biol* **24**: 1429–34
- 736 Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F,
737 Nieto MA (2000) The transcription factor snail controls epithelial-mesenchymal transi-
738 tions by repressing E-cadherin expression. *Nat Cell Biol* **2**: 76–83
- 739 Chang H, Park C, Parvin B (2007) Quantitative Representation of Three-dimensional Cell
740 Culture Models. In *Proceedings of the 2007 IEEE International Symposium on Biomed-*
741 *ical 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-

- 769 duction. *J Leukoc Biol* **71**: 731–40
- 770 Derynck R, Zhang YE (2003) Smad-dependent and Smad-independent pathways in TGF-
771 beta family signalling. *Nature* **425**: 577–84
- 772 Dhasarathy A, Phadke D, Mav D, Shah RR, Wade PA (2011) The transcription factors
773 Snail and Slug activate the transforming growth factor-beta signaling pathway in breast
774 cancer. *PLoS One* **6**: e26514
- 775 Dhimolea E, Maffini MV, Soto AM, Sonnenschein C (2010) The role of collagen reorgani-
776 zation on mammary epithelial morphogenesis in a 3D culture model. *Biomaterials* **31**:
777 3622–3630
- 778 Eastman Q, Grosschedl R (1999) Regulation of LEF-1/TCF transcription factors by Wnt
779 and other signals. *Curr Opin Cell Biol* **11**: 233–40
- 780 Eisenberg LM, Markwald RR (1995) Molecular regulation of atrioventricular valvuloseptal
781 morphogenesis. *Circ Res* **77**: 1–6
- 782 Fan JM, Ng YY, Hill PA, Nikolic-Paterson DJ, Mu W, Atkins RC, Lan HY (1999) Trans-
783 forming growth factor-beta regulates tubular epithelial-myofibroblast transdifferentiation
784 in vitro. *Kidney Int* **56**: 1455–1467
- 785 Ferrara N (2002) VEGF and the quest for tumour angiogenesis factors. *Nat Rev Cancer*
786 **2**: 795–803
- 787 Fonseca C, Fleming P, *et al.* (1993) Genetic algorithms for multiobjective optimization:
788 Formulation, discussion and generalization. In *Proceedings of the fifth international*
789 *conference on genetic algorithms*, vol. 423. Citeseer, pp. 416–423
- 790 Gadkar KG, Doyle FJ 3rd, Crowley TJ, Varner JD (2003) Cybernetic model predictive
791 control of a continuous bioreactor with cell recycle. *Biotechnol Prog* **19**: 1487–97
- 792 Gadkar KG, Varner J, Doyle FJ 3rd (2005) Model identification of signal transduction
793 networks from data using a state regulator problem. *Syst Biol (Stevenage)* **2**: 17–30
- 794 Gatza CE, Oh SY, Blobe GC (2010) Roles for the type III TGF-beta receptor in human
795 cancer. *Cell Signal* **22**: 1163–74

- 796 Grumolato L, Liu G, Haremake T, Mungamuri SK, Mong P, Akiri G, Lopez-Bergami P, Arita
797 A, Anouar Y, Mlodzik M, Ronai ZA, Brody J, Weinstein DC, Aaronson SA (2013) ?-
798 Catenin-Independent Activation of TCF1/LEF1 in Human Hematopoietic Tumor Cells
799 through Interaction with ATF2 Transcription Factors. *PLoS Genet* **9**: e1003603
- 800 Guaita S, Puig I, Franci C, Garrido M, Dominguez D, Batlle E, Sancho E, Dedhar S,
801 De Herreros AG, Baulida J (2002) Snail induction of epithelial to mesenchymal transition
802 in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *J Biol Chem*
803 **277**: 39209–16
- 804 Hales AM, Schulz MW, Chamberlain CG, McAvoy JW (1994) TGF-beta 1 induces lens
805 cells to accumulate alpha-smooth muscle actin, a marker for subcapsular cataracts.
806 *Curr Eye Res* **13**: 885–890
- 807 Handl J, Kell DB, Knowles J (2007) Multiobjective optimization in bioinformatics and com-
808 putational biology. *IEEE/ACM Trans Comput Biol Bioinform* **4**: 279–92
- 809 Hardy KM, Booth BW, Hendrix MJC, Salomon DS, Strizzi L (2010) ErbB/EGF signaling
810 and EMT in mammary development and breast cancer. *J Mammary Gland Biol Neo-*
811 *plasia* **15**: 191–9
- 812 Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, Allgöwer F (2011) Iden-
813 tification of models of heterogeneous cell populations from population snapshot data.
814 *BMC Bioinformatics* **12**: 125
- 815 Hemavathy K, Ashraf SI, Ip YT (2000a) Snail/slugg family of repressors: slowly going into
816 the fast lane of development and cancer. *Gene* **257**: 1–12
- 817 Hemavathy K, Guru SC, Harris J, Chen JD, Ip YT (2000b) Human Slug is a repressor that
818 localizes to sites of active transcription. *Mol Cell Biol* **20**: 5087–95
- 819 Hindmarsh A, Brown P, Grant K, Lee S, Serban R, Shumaker D, Woodward C (2005) SUN-
820 DIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions*
821 *on Mathematical Software* **31**: 363–396
- 822 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

- 850 alveolar epithelial to mesenchymal cell transition (EMT). *Respir Res* **6**: 56
- 851 Kavsak P, Rasmussen RK, Causing CG, Bonni S, Zhu H, Thomsen GH, Wrana JL (2000)
- 852 Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor
- 853 for degradation. *Mol Cell* **6**: 1365–1375
- 854 Kim K, Lu Z, Hay ED (2002) Direct evidence for a role of beta-catenin/LEF-1 signaling
- 855 pathway in induction of EMT. *Cell Biol Int* **26**: 463–76
- 856 Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and
- 857 model selection. In *International joint Conference on artificial intelligence*, vol. 14. Cite-
- 858 seer, pp. 1137–1145
- 859 Kretschmar M, Doody J, Timokhina I, Massagué J (1999) A mechanism of repression of
- 860 TGFbeta/ Smad signaling by oncogenic Ras. *Genes Dev* **13**: 804–16
- 861 Larue L, Bellacosa A (2005) Epithelial-mesenchymal transition in development and can-
- 862 cer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene* **24**: 7443–54
- 863 Lee MW, Vassiliadis VS, Park JM (2009) Individual-based and stochastic modeling of cell
- 864 population dynamics considering substrate dependency. *Biotechnol Bioeng* **103**: 891–9
- 865 Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and anal-
- 866 ysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol* **7**: e1002263
- 867 Lian Yg, Zhou Qg, Zhang Yj, Zheng FI (2011) VEGF ameliorates tubulointerstitial fibrosis
- 868 in unilateral ureteral obstruction mice via inhibition of epithelial-mesenchymal transition.
- 869 *Acta Pharmacol Sin* **32**: 1513–21
- 870 Linding R, Jensen LJ, Ostheimer GJ, van Vugt MATM, Jørgensen C, Miron IM, Diella F,
- 871 Colwill K, Taylor L, Elder K, Metalnikov P, Nguyen V, Pasculescu A, Jin J, Park JG, Sam-
- 872 son LD, Woodgett JR, Russell RB, Bork P, Yaffe MB, *et al.* (2007) Systematic discovery
- 873 of in vivo phosphorylation networks. *Cell* **129**: 1415–26
- 874 Lopez-Casillas F, Riquelme C, Perez-Kato Y, Ponce-Castaneda MV, Osses N, Esparza-
- 875 Lopez J, Gonzalez-Nunez G, Cabello-Verrugio C, Mendoza V, Troncoso V, Brandan
- 876 E (2003) Betaglycan expression is transcriptionally up-regulated during skeletal mus-

877 cle differentiation. Cloning of murine betaglycan gene promoter and its modulation by
878 MyoD, retinoic acid, and transforming growth factor-beta. *J Biol Chem* **278**: 382–390

879 Luan D, Szlam F, Tanaka KA, Barie PS, Varner JD (2010) Ensembles of uncertain mathe-
880 matical models can identify network response to therapeutic interventions. *Mol Biosyst*
881 **6**: 2272–86

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

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

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

888 Massagué J, Seoane J, Wotton D (2005) Smad transcription factors. *Genes Dev* **19**:
889 2783–810

890 Masszi A, Di Ciano C, Sirokmány G, Arthur WT, Rotstein OD, Wang J, McCulloch CAG,
891 Rosivall L, Mucsi I, Kapus A (2003) Central role for Rho in TGF-beta1-induced alpha-
892 smooth muscle actin expression during epithelial-mesenchymal transition. *Am J Physiol*
893 *Renal Physiol* **284**: F911–24

894 Medici D, Hay ED, Goodenough DA (2006) Cooperation between snail and LEF-1 tran-
895 scription factors is essential for TGF-beta1-induced epithelial-mesenchymal transition.
896 *Mol Biol Cell* **17**: 1871–9

897 Medici D, Hay ED, Olsen BR (2008) Snail and Slug promote epithelial-mesenchymal tran-
898 sition through beta-catenin-T-cell factor-4-dependent expression of transforming growth
899 factor-beta3. *Mol Biol Cell* **19**: 4875–87

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

903 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.

- 931 *Dev Cell* **7**: 21–32
- 932 Oft M, Peli J, Rudaz C, Schwarz H, Beug H, Reichmann E (1996) TGF-beta1 and Ha-Ras
933 collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor
934 cells. *Genes Dev* **10**: 2462–2477
- 935 Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K (2010) Heterogeneity for stem
936 cell-related markers according to tumor subtype and histologic stage in breast cancer.
937 *Clin Cancer Res* **16**: 876–87
- 938 Pearson GW, Hunter T (2007) Real-time imaging reveals that noninvasive mammary ep-
939 ithelial acini can contain motile cells. *J Cell Biol* **179**: 1555–67
- 940 Peinado H, Portillo F, Cano A (2004) Transcriptional regulation of cadherins during devel-
941 opment and carcinogenesis. *Int J Dev Biol* **48**: 365–75
- 942 Peinado H, Quintanilla M, Cano A (2003) Transforming growth factor beta-1 induces snail
943 transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal tran-
944 sitions. *J Biol Chem* **278**: 21113–21123
- 945 Phanish MK, Wahab NA, Colville-Nash P, Hendry BM, Dockrell MEC (2006) The differen-
946 tial role of Smad2 and Smad3 in the regulation of pro-fibrotic TGFbeta1 responses in
947 human proximal-tubule epithelial cells. *Biochem J* **393**: 601–607
- 948 Polizzotti L, Basak O, Bjornsson C, Shubert K, Yener B, Plopper G (2012) Novel Image
949 Analysis Approach Quantifies Morphological Characteristics of 3D Breast Culture Acini
950 with Varying Metastatic Potentials. *J Biomed Biotech* **2012**: 1–16
- 951 Polyak K, Weinberg RA (2009) Transitions between epithelial and mesenchymal states:
952 acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**: 265–273
- 953 Rodriguez-Fernandez M, Rehberg M, Kremling A, Banga JR (2013) Simultaneous model
954 discrimination and parameter estimation in dynamic models of cellular systems. *BMC*
955 *Syst Biol* **7**: 76
- 956 Sainani KL (2012) Meet the Skeptics: Why some doubt biomedical models - and what it
957 takes to win them over. *Biomedical Computation Review* : 12 – 18

- 958 Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of
959 the dynamics of the MAP kinase cascade activated by surface and internalized EGF
960 receptors. *Nat Biotechnol* **20**: 370–375
- 961 Singh G, Singh SK, König A, Reutlinger K, Nye MD, Adhikary T, Eilers M, Gress TM,
962 Fernandez-Zapico ME, Ellenrieder V (2010) Sequential activation of NFAT and c-Myc
963 transcription factors mediates the TGF-beta switch from a suppressor to a promoter of
964 cancer cell proliferation. *J Biol Chem* **285**: 27241–50
- 965 Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models
966 using Pareto Optimal Ensemble Techniques (POETs). *Biotechnol J* **5**: 768–80
- 967 Stahl PJ, Felsen D (2001) Transforming growth factor-beta, basement membrane, and
968 epithelial-mesenchymal transdifferentiation: implications for fibrosis in kidney disease.
969 *Am J Pathol* **159**: 1187–92
- 970 Strauss R, Li ZY, Liu Y, Beyer I, Persson J, Sova P, Möller T, Pesonen S, Hemminki A,
971 Hamerlik P, Drescher C, Urban N, Bartek J, Lieber A (2011) Analysis of epithelial and
972 mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plastic-
973 ity. *PLoS One* **6**: e16186
- 974 Sullivan NJ, Sasser AK, Axel AE, Vesuna F, Raman V, Ramirez N, Oberyszyn TM, Hall BM
975 (2009) Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human
976 breast cancer cells. *Oncogene* **28**: 2940–7
- 977 Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochas-
978 ticity in gene expression. *Proc Natl Acad Sci U S A* **99**: 12795–800
- 979 Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, Varner JD (2010) Analysis of the molec-
980 ular networks in androgen dependent and independent prostate cancer revealed fragile
981 and robust subsystems. *PLoS One* **5**: e8864
- 982 Tasseff R, Nayak S, Song SO, Yen A, Varner JD (2011) Modeling and analysis of retinoic
983 acid induced differentiation of uncommitted precursor cells. *Integr Biol (Camb)* **3**: 578–
984 591

- 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, Böttinger 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

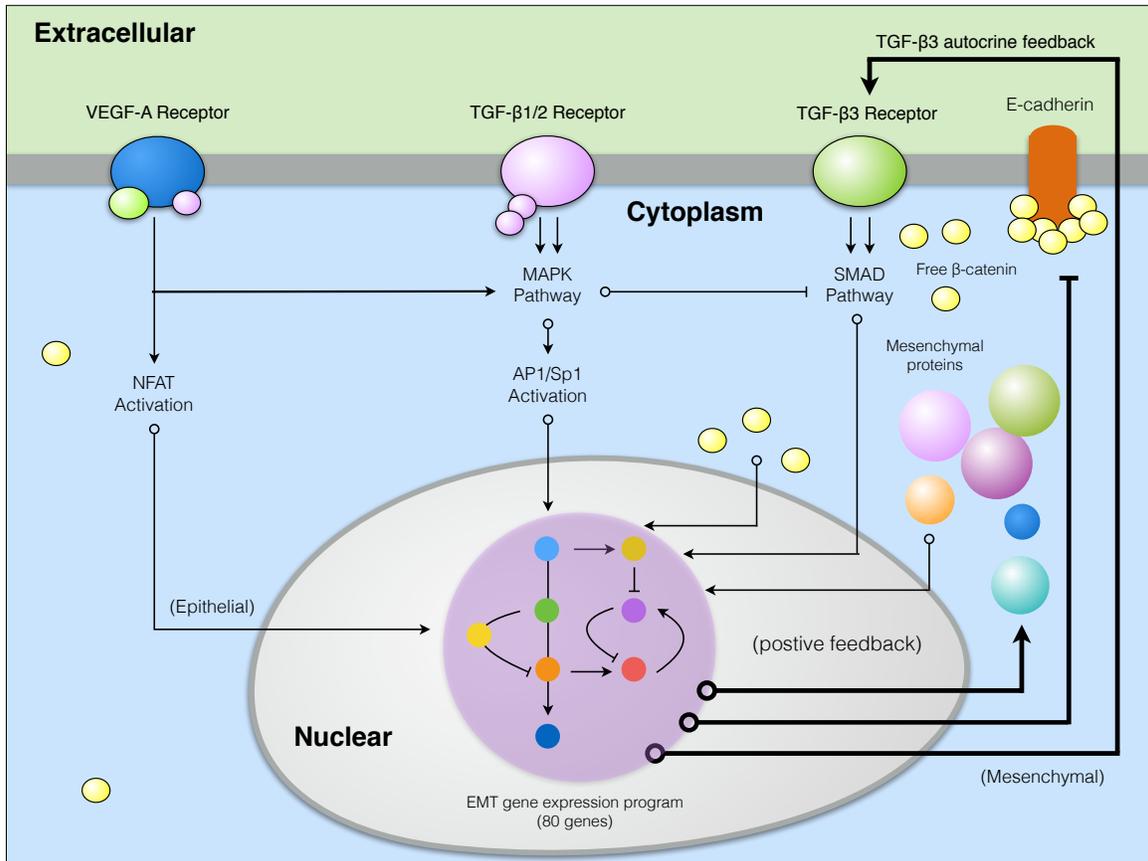


Fig. 1: Model connectivity recreates the core architecture during EMT. The EMT network contains 995 nodes (proteins, mRNA, and genes) interconnected by 1700 interactions. Central to EMT induction, activation of the MAPK cascade occurs through TGF- β 1/2 binding which activates the AP-1/Sp1 transcriptional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

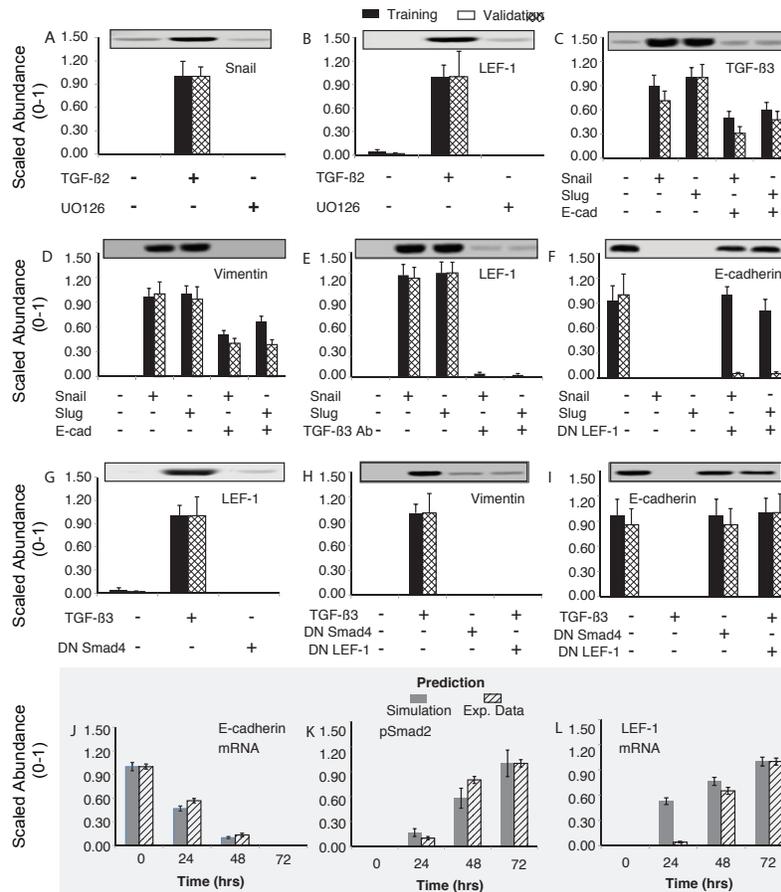


Fig. 2: Training and validation simulations. The population of EMT models qualitatively captured TGF β -induced EMT signaling. (A-I) The population was generated using POETs and trained using 11 different objective functions (41 data sets) taken from Medici *et al.* Medici *et al.* (2008). The model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

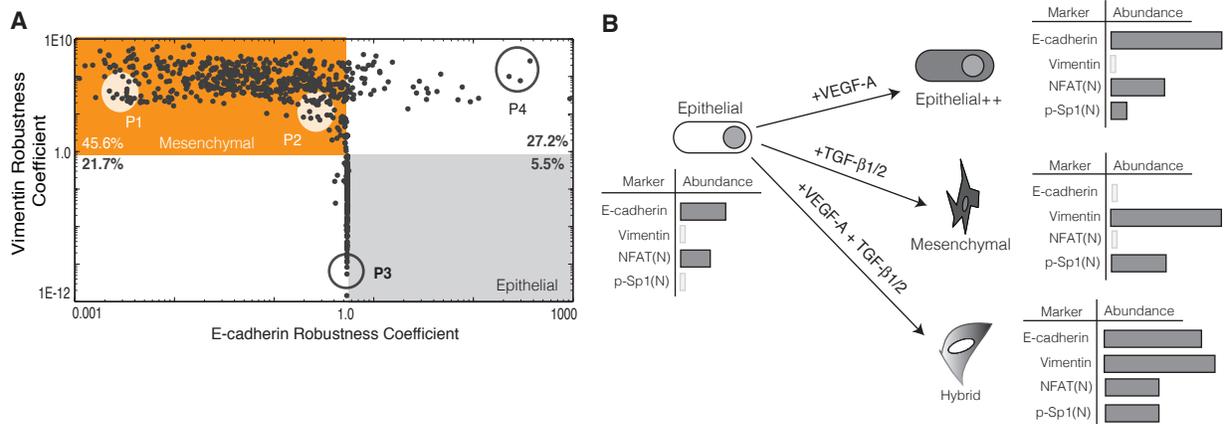


Fig. 3: Simulated $TGF\beta1/2$ exposure promoted phenotype heterogeneity. Robustness coefficients were used to quantify the effect of perturbations after steady state across the population of data sets. Coefficients with values > 1 (< 1) indicated a marker increased (decreased) compared to a base state, while a value of 1 indicated approximately no change following a perturbation. E-cadherin and Vimentin robustness coefficients were used as phenotypic markers. (A) $TGF\beta1/2$ perturbation, we isolated 4 distinct parameter sets known to have phenotypically different behaviors. (B) Each region has been represented by a small signaling network. From our analysis, we determined that the differences were a function of downstream transcription factors (phosphorylated-Sp1, and NFAT) within the $TGF\beta2$ and VEGFA pathway, respectively. (C) We hypothesized that elevated phosphorylated-Sp1 and NFAT levels could drive phenotype heterogeneity through simultaneous $TGF\beta2$ and VEGFA treatment, similar to region three (R3).

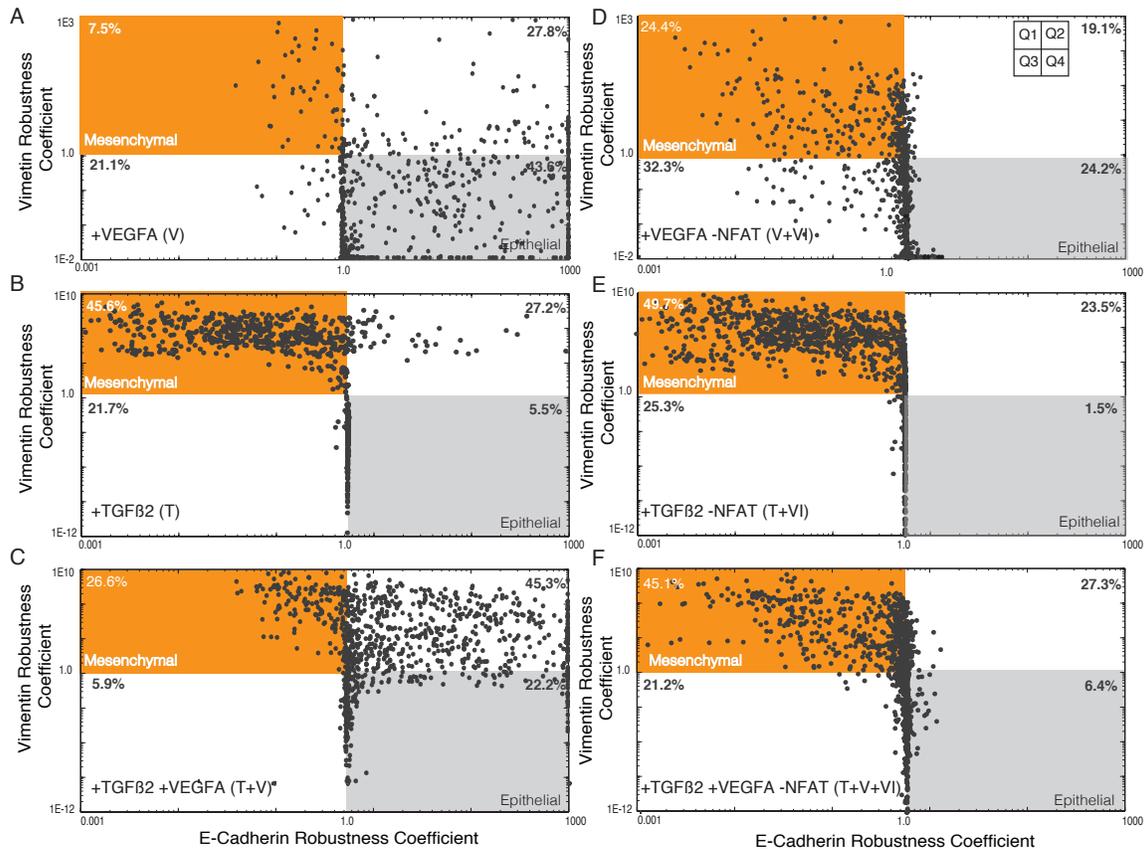


Fig. 4: Simulated response to $TGF\beta 1/2$ and VEGF-A exposure with and without axis specific inhibitors. Robustness coefficients were used to quantify the shift in population at 48 hrs. (A-C) VEGF-A (50 a.u.) treatment resulted in a population with enhanced epithelial (Q4) properties. This was contrary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

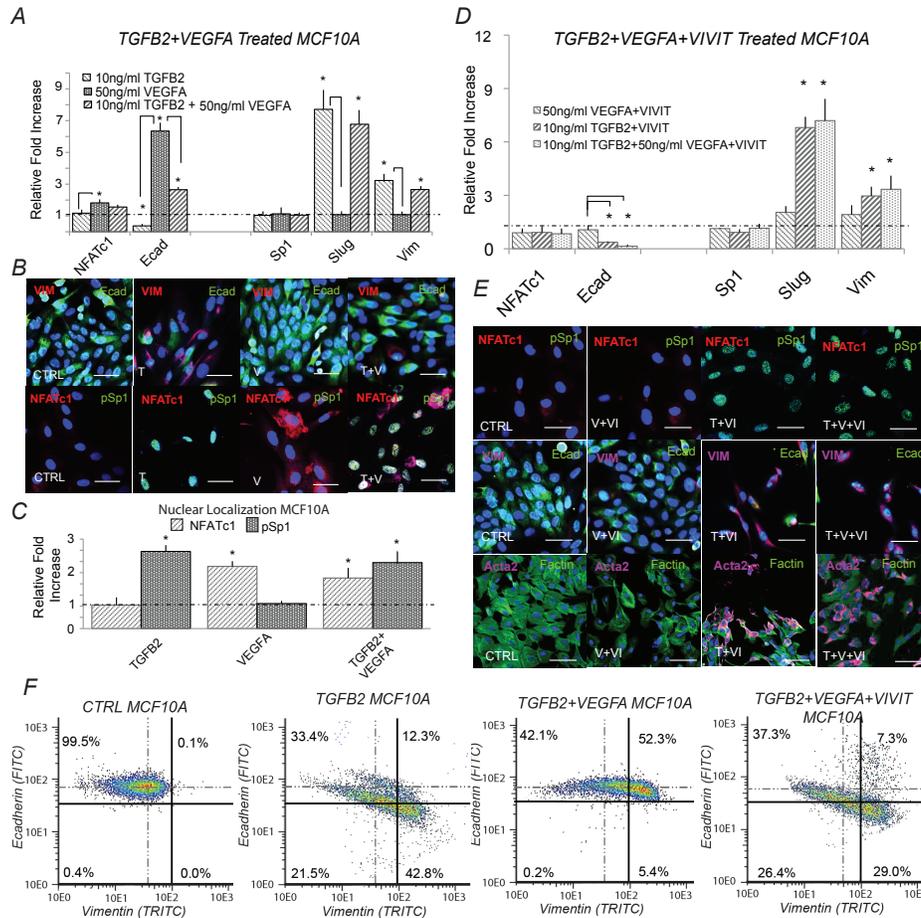


Fig. 5: Simultaneous TGF- β 1/2 and VEGF-A treatment induced phenotype heterogeneity and is dependent upon NFAT activity *in-vitro*. (A) In MCF10A, treatment with (10ng/ml) TGF β 2 increased Slug and vimentin, while ecadherin expression was inhibited at both the gene and protein level at 48 hrs. Conversely,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

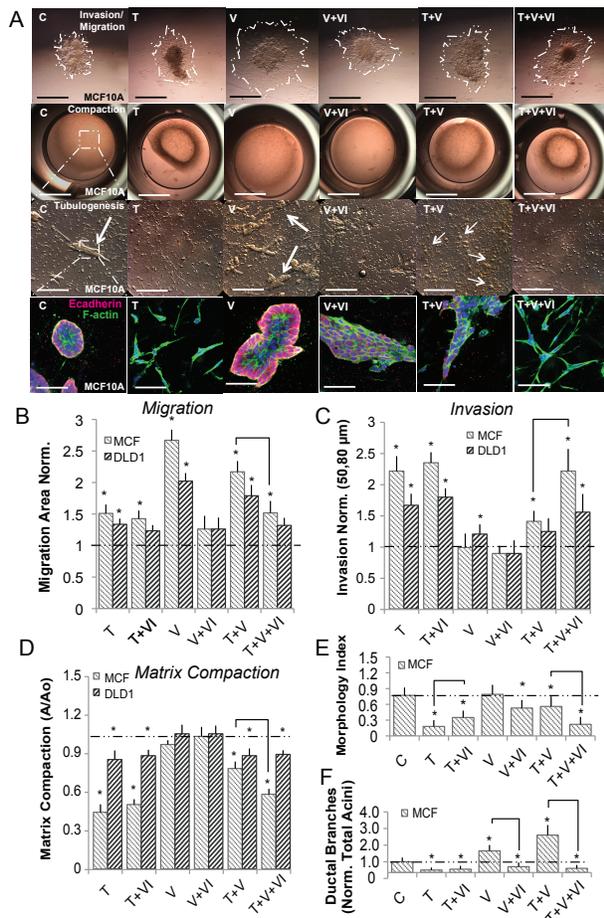


Fig. 6: Ductal branching is dependent upon phenotype heterogeneity within MCF10A in 3-D culture. MCF10A and DLD1 were formed into spheroids overnight and explanted to a collagen gel for 72 hrs. For compaction and tubular assays, cells were embedded into collagen gels for 72 hrs, and the extent of tubulogenesis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

1038 **Supplemental Materials and Methods**

1039 **EMT model network architecture.** TGF β is a major inducer of EMT in development, fi-
1040 brosis, and carcinogenesis with different isoforms mediating various effects depending on
1041 specific cellular context Nawshad *et al.* (2005). TGF β 1 was first described as an inducer
1042 of EMT in normal mammary epithelial cells Miettinen *et al.* (1994) and has since been
1043 shown to mediate EMT in vitro in a number of different epithelial cells, including renal
1044 proximal tubular, DLD1 colon carcinoma, and most recently alveolar epithelial cells Fan
1045 *et al.* (1999), Hales *et al.* (1994), Kasai *et al.* (2005), Willis *et al.* (2005). TGF β signaling
1046 occurs through the Smad pathway in which signals are transduced by transmembrane
1047 serine/threonine kinase type I (ALK5) and type II (TGF β RII) receptors. To increase lig-
1048 and affinity, betaglycan (TGF β RIII) can also interact with TGF β RI,II Gatza *et al.* (2010).
1049 Upon TGF β stimulation, the receptors are internalized into early endosomes where Smad
1050 anchor for receptor activation (SARA) modulates formation of complexes with (R-Smad)
1051 Smad2 or Smad3. Smad2 and Smad3 are then phosphorylated at serine residues by
1052 the type I receptor Massagué *et al.* (2005). Phosphorylation induces their association
1053 with (Co-Smad) Smad4 and translocation to the nucleus where they interact with other
1054 transcription factors to regulate the transcription of TGF β responsive genes, including
1055 alpha-smooth muscle actin, collagen1A2, vimentin, fibronectin, and plasminogen activa-
1056 tor inhibitor-1 (PAI-1) by interacting with Smad-binding elements Dennler *et al.* (2002),
1057 Derynck & Zhang (2003). To regulate TGF β signaling, Smurf2 (a ubiquitin E3 ligase) can
1058 become activated to mediate proteasome dependent degradation of Smad2 or bind with
1059 Smad 7 to target TGF β receptor for degradation Bonni *et al.* (2001), Kavsak *et al.* (2000).
1060 *Cell Type Dependency* Interestingly, differential roles for Smad2 and Smad3 in TGF β
1061 induced EMT have been demonstrated. For example, using primary cells from mice
1062 with hepatocyte-specific double knockout of Smad2 and Smad3, it was demonstrated
1063 that Smad3 but not Smad2 was required for a key morphological changes and induc-
1064 tion of EMT Ju *et al.* (2006). In contrast, using human proximal tubular epithelial cells,

1065 increased colony growth factor and decreased E-cadherin were Smad3 dependent, in-
1066 creased MMP-2 was Smad2 dependent, while alpha-smooth muscle actin was dependent
1067 on both Phanish *et al.* (2006). Furthermore, Smad2 signaling has been demonstrated in
1068 cancerous lung epithelial A549 cells Kasai *et al.* (2005). Inhibition of Smad3 signaling de-
1069 creases the metastatic potential of xenografted breast cancer cell lines Tian *et al.* (2004,
1070 2003). Together, these results suggest that the precise Smad pathway activated may de-
1071 pend on the particular cellular context. Regardless, a recent transcriptomic analysis of
1072 TGF β induced EMT in normal mouse and human epithelial cells using a dominant neg-
1073 ative approach demonstrated that Smad signaling was critical for regulation of all tested
1074 target genes Valcourt *et al.* (2005).

1075 *Smad-Independent EMT Induction Pathways.* Although less well established than the
1076 Smad-dependent pathways in the induction of EMT, there is substantial evidence for
1077 TGF β activation of Smad-independent signaling in some aspects of this process. These
1078 include RhoA, MAPK, PI3 kinase, and Notch signaling pathway, which have mostly been
1079 studied in-vitro. For example, the small GTPase RhoA is involved in TGF β induced EMT
1080 in a number of cell types including NMuMG mammary epithelial cells and mink lung ep-
1081 ithelial (Mv1Lu) cells particularly in the regulation of cytoskeletal and adherens junction
1082 rearrangement. In addition to its role in cytoskeletal remodeling, Rho has been shown to
1083 activate the alpha-smooth muscle promoter during EMT in kidney proximal tubular (LL-
1084 CPK1) cells Masszi *et al.* (2003). TGF β has been shown to elevate ERK activity in culture
1085 models of EMT (human keratinocytes, NMuMG mammary epithelial cells, and mouse
1086 cortical tubule epithelial cells) Davies *et al.* (2005), Xie *et al.* (2004), Zavadil *et al.* (2001).
1087 This ERK activity was required for disassembly of adherens junctions and induction of cell
1088 motility Xie *et al.* (2004). Inhibition of ERK blocked these key morphological changes of
1089 EMT in mammary gland epithelial cells Xie *et al.* (2003). Recent studies in MDCKII and
1090 DLD1 colon carcinoma cells revealed that TGF β 3 can signal through PI3 kinase. Upon
1091 activation, PI3 kinase activates molecules such as ILK and AKT, which can phosphory-

1092 late and inactivate GSK3 β , a protein that targets both Snail and β -catenin for degradation
1093 through the ubiquitin proteasome pathway Medici *et al.* (2008). Lastly, there is increasing
1094 evidence for a role of Notch pathways in regulating EMT, particularly during development
1095 Nawshad *et al.* (2005). In cardiac cushion EMT, it was found that that Slug is directly
1096 up-regulated by Notch in endothelial cells repression of the vascular endothelial cadherin
1097 promoter and for promoting migration of transformed endothelial cells. In whole, the dis-
1098 tinction between Smad-dependent and Smad-independent mechanisms remains difficult
1099 to interpret due to the significant cross talk between these pathways. In most cases, stim-
1100 ulation of these cooperative pathways provides the context for induction and specification
1101 of EMT within a particular tissue/cell type, with Smads representing the dominant path-
1102 way, which in some instances may be necessary but not sufficient for induction of full EMT
1103 Zavadil & Böttinger (2005).

1104 *Master Transcription Regulators* Master transcription factors of EMT are thought to be
1105 regulated by Snail1 (SNAI1) and Snail2 (SNAI2) (previously known as Snail and Slug,
1106 respectively). These transcription factors are zinc finger proteins that function as repres-
1107 sors of E-cadherin transcription in cultured epithelial cells Cano *et al.* (2000), repression
1108 of E-cadherin leads to dissolution of adherens junctions. Snail and Slug can be activated
1109 by TGF β via both Smad-dependent and -independent pathways in a cell type dependent
1110 fashion in cultured cells Peinado *et al.* (2003). Differential expression of Snail and Slug
1111 is observed in TGF β induced EMT in keratinocytes, renal proximal tubular, and mam-
1112 mary epithelial cells, suggesting that they are regulated in a cell specific. Interestingly,
1113 key regulatory units of Snail and Slug appear to be regulated by upstream transcription
1114 factors AP1 and SP1, respectively. Peinado *et al.* clearly established a necessary role
1115 of Snail in mediating TGF β 1 induced EMT in MDCK cells. They demonstrated that the
1116 Ras-Raf-MEK-ERK-AP1 signaling pathway could up-regulate synthesis of the E-cadherin
1117 repressor molecule Snail (directly binds to Snail promoter) Peinado *et al.* (2004). Like-
1118 wise, TGF β 1 induced the expression of Slug in both lens and other epithelial cells in vitro.

1119 The Sp1 binding site in the Slug promoter is largely responsible for TGF β 1 induced Slug
1120 expression and upstream of MAPK signaling. In addition, the TGF β 1 mediated repression
1121 of E-cadherin was significantly inhibited by Slug siRNA Choi *et al.* (2007).

1122 *ERK/MAPK Crosstalk* The ERK-MAPK pathway has been shown to modify TGF β sig-
1123 naling at multiple levels. One way is that the activated Ras pathway inhibits the TGF β
1124 induced nuclear accumulation of Smad2/3, as seen in epithelial cells Zhang (2009). ERK
1125 kinases have been shown to phosphorylate Smad2 and Smad3 at specific sites in the re-
1126 gion linking the MH1 and MH2 domains. These sites are separate from the TGF β receptor
1127 phosphorylation sites which activate Smad nuclear translocation Massague (2003). The
1128 effect of interaction between ERKs and Smads is the subject of some controversy, with
1129 data suggesting that such an interaction either enhances or inhibits downstream events.
1130 A hyperactive Ras pathway has been shown to effectively counteract the antiproliferative
1131 activity of TGF β through attenuation of Smad accumulation in the nucleus Kretzschmar
1132 *et al.* (1999). In contrast, Ras signals strongly cooperated with Smads for invasion of
1133 human carcinoma cells Oft *et al.* (1996). A second way is through transcriptional regula-
1134 tion. Recently, it was found that Sp1, via transcriptional induction of Vimentin, cooperates
1135 with activated Smad complexes in mesenchymal transition and migration of pancreatic
1136 cancer cells upon TGF β stimulation Jungert *et al.* (2007). Likewise, binding sites at the
1137 SP1/Smad3 complex was found to regulate the betaglycan receptors Lopez-Casillas *et al.*
1138 (2003). Taken together, multiple levels of cross-talk (both positive and negative) exist
1139 within MAPK and TGF β , and may also be cell type dependent.

1140 **Analysis of the signal flow through the EMT architecture.** To investigate temporal
1141 shifts for key species dominating the EMT response, we calculated the scaled flux through
1142 the signaling architecture (Fig. S1). Three modes of operation were simulated to iden-
1143 tify distinct behavioral differences: (a) no TGF- β 2 (-TGF- β 1/2), (b) TGF- β 1/2 stimulation
1144 (+TGF- β 2), and (c) TGF- β 1/2 stimulation while blocking the autocrine response of TGF-
1145 β 3 (+TGF- β 1/2 + TGF- β 3Ab). Following TGF- β 1/2 stimulation, AP1/SP1 rapidly stimu-

1146 lates Snail/Slug expression (Fig. S1A) within 1 hr. During this phase, Snail/Slug repress
1147 E-cadherin expression, which in turn reduces the interaction between E-cadherin and
1148 β -catenin, and the accumulation of free cytosolic β -catenin ensues. Conversely, in the
1149 presences of a TGF- β 3 neutralizing antibody, membrane bound β -catenin increased af-
1150 ter 10 hrs, while low levels of free cytosolic β -catenin accumulated between 1-5 hrs (Fig.
1151 S2A,B). Accumulation of free β -catenin is critical for complexing with TCF4 to rapidly
1152 produce the autocrine response of TGF- β 3 within 1-10 hrs. Following TGF- β 3 expres-
1153 sion, formation and spatial relocation of the Smad complexes occurs in a time dependent
1154 manner. Smad2 is phosphorylated within 1 hr and nuclear localization of the pSmad2/4
1155 complex dramatically increases after 10 hours (Fig. S1C). This timely formation is crit-
1156 ical for completion of the EMT process. At steady state, the E-cadherin complex was
1157 maintained at a basal epithelial level with no expression of Vimentin (Fig. S1D). During
1158 TGF- β 1/2 stimulation, both MAPK and Smad act synergistically to repress the E-cadherin
1159 complex within 1 hr. This is followed by increase of Vimentin at \sim 1hr, while exponentially
1160 increasing around 10 hrs. Both species elevate within 1 hr, however the complex does not
1161 significantly form until \sim 10 hrs. When blocking TGF- β 3, Snail/Slug downregulates the
1162 E-cadherin complex between 5-10 hrs, but cannot complete the transformation due to the
1163 loss of Smad2 phosphorylation (Fig. S1D).

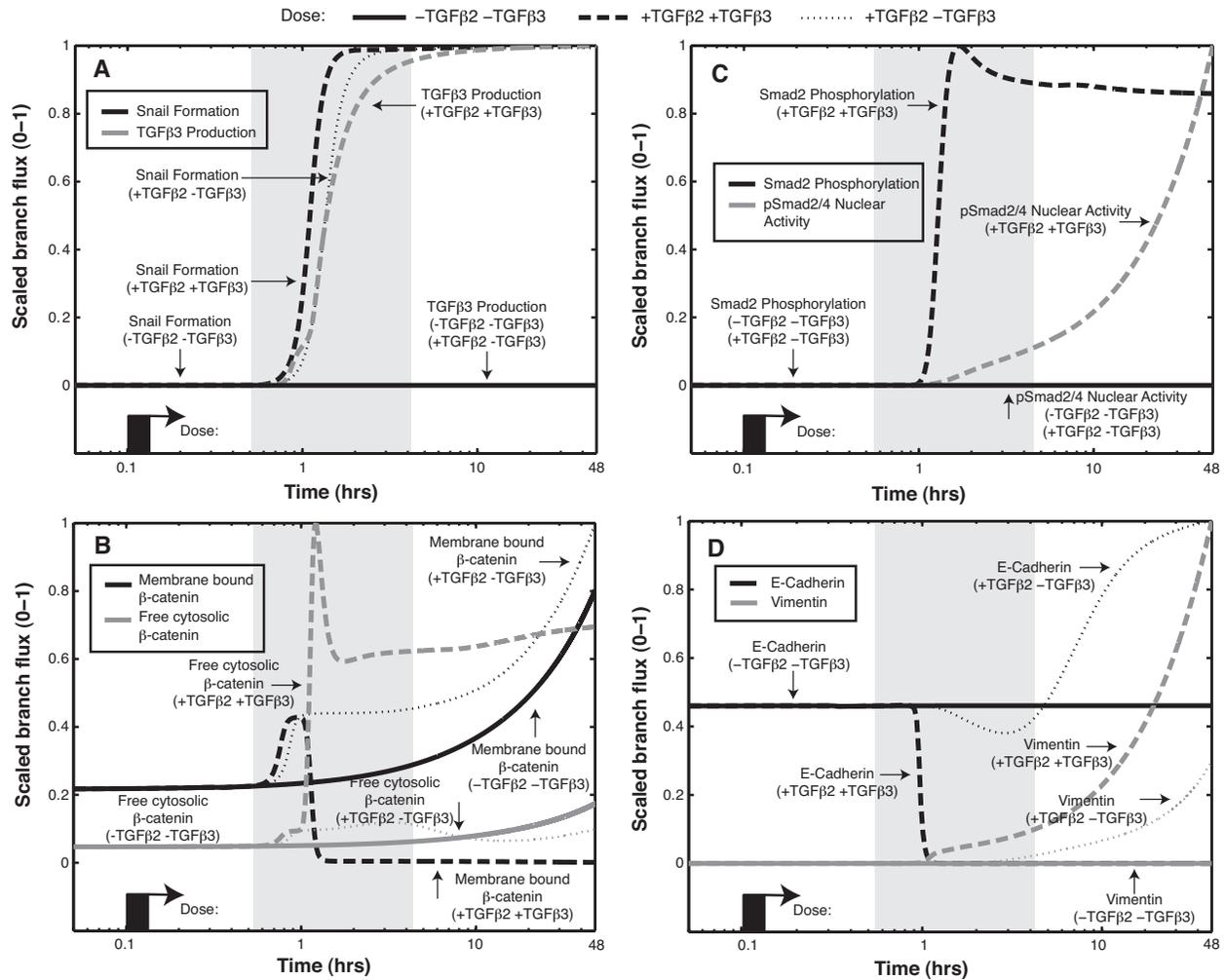


Fig. S1: Signal flow analysis of key species at steady state, TGF-β1/2 stimulation, and blocking the TGF-β3 autocrine response. (A) The MAPK cascade is directly responsible for rapid expression of Snail and downstream TGF-β3 formation (1 hrs). (B) TGF-β1/2 reduces β-catenin, allowing rapid free-cytosolic 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.

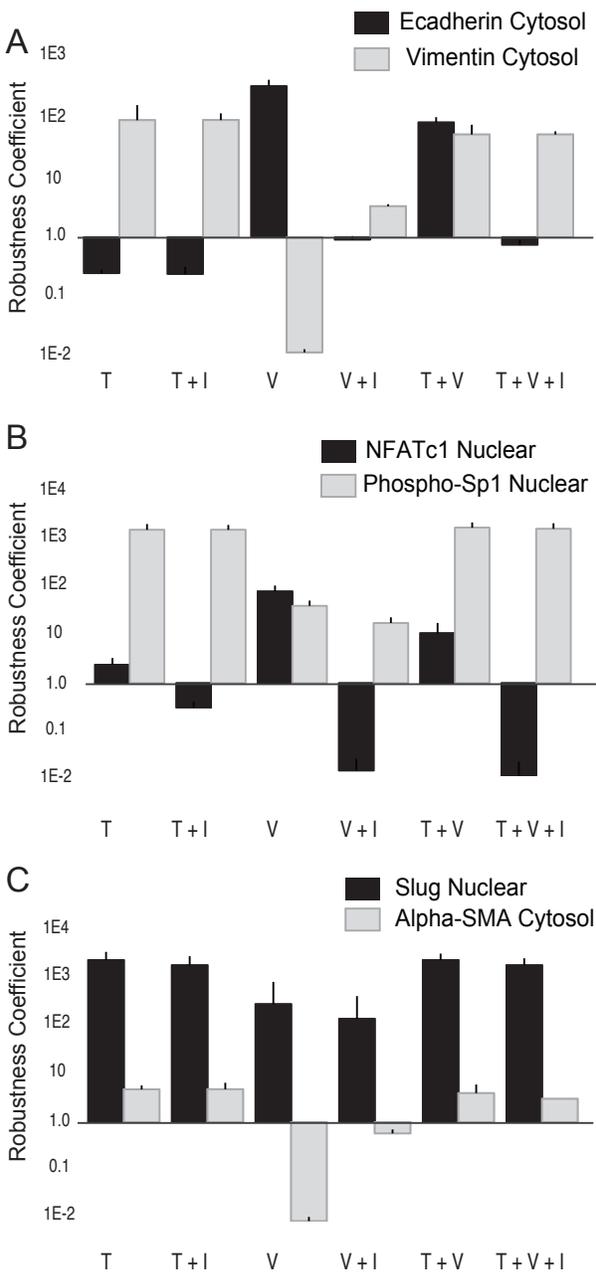


Fig. S2: Robustness analysis for key molecular species at $t = 48$ hrs for combinations of TGF- β 1/2, VEGF-A and NFATc1 inhibitors. Robustness coefficients for the indicated species were calculated for $N \sim 1100$ ensemble members for 48 hrs following the addition of TGF- β 1/2 (T), TGF- β 1/2 + NFATc1 inhibitor (T + I), VEGF-A (V), VEGF-A + NFATc1 inhibitor (V + I) and TGF- β 1/2 + VEGF-A (T+V) + NFATc1 inhibitor (T + V + I). (A) Robustness coefficients for E-cadherin and Vimentin as a function of condition. (B) Robustness coefficients for nuclear localized phosphorylated Sp1 and NFATc1 as a function of condition. (C) Robustness coefficients for nuclear localized Slug and 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=TGF β 2, V=VEGFA, VI= NFAT inhibitor (VIVIT).

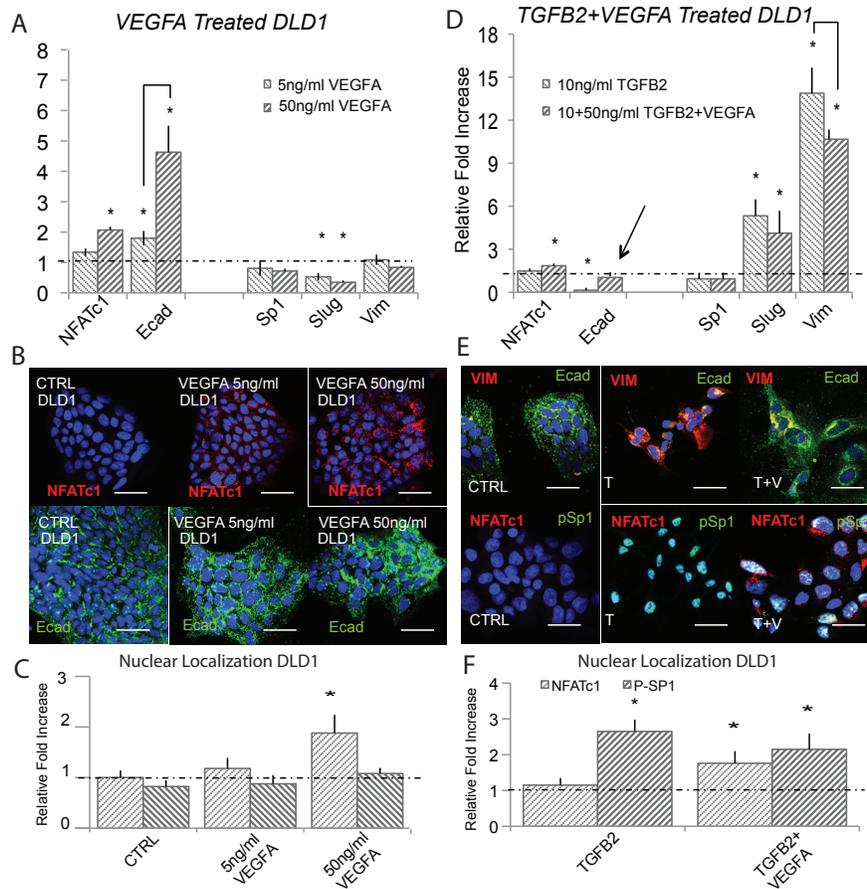


Fig. S3: VEGF-A attenuates TGF- β 1/2 to induce phenotype heterogeneity in DLD1. (A) In DLD1, we found that 5ng/ml of VEGFA increased NFATc1 and E-cadherin gene expression via qPCR and 50ng/ml potentiated this effect at 48 hrs. (B - C) These findings were confirmed at the protein level via immunofluorescence,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

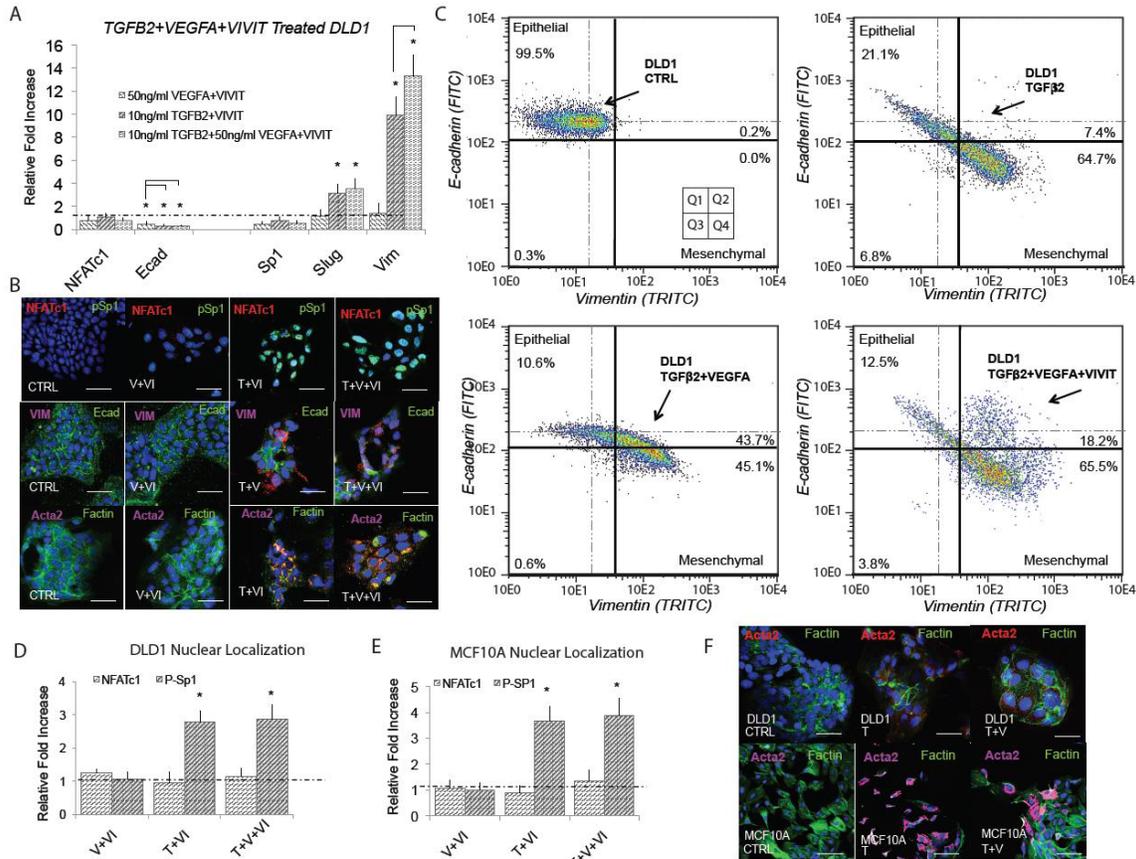


Fig. S4: E-cadherin expression is dependent upon NFAT activity in DLD1. (A) Treatment with VEGFA (50ng/ml) and NFAT inhibitory peptide VIVIT (10 μ M) resulted in significantly reduced ecadherin expression (qRT-PCR at 48hrs). Addition of TGF β 2 (10ng/ml) and VIVIT resulted in increased Slug and vimentin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

1164 **Estimation and cross-validation of EMT model parameters.** We used the Pareto
1165 Optimal Ensemble Technique (POETs) multiobjective optimization framework in combi-
1166 nation with leave-one-out cross-validation to estimate an ensemble of TGF β /EMT mod-
1167 els. Cross-validation was used to calculate both training and prediction error during the
1168 parameter estimation procedure Kohavi (1995). The 41 intracellular protein and mRNA
1169 data-sets used for identification were organized into 11 objective functions. These 11
1170 objective functions were then partitioned, where each partition contained ten training ob-
1171 jectives and one validation objective. POETs integrates standard search strategies e.g.,
1172 Simulated Annealing (SA) or Pattern Search (PS) with a Pareto-rank fitness assignment
1173 Song *et al.* (2010). Denote a candidate parameter set at iteration $i + 1$ as \mathbf{k}_{i+1} . The
1174 squared error for \mathbf{k}_{i+1} for training set j was defined as:

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

1175 The symbol $\hat{\mathcal{M}}_{ij}$ denotes scaled experimental observations (from training set j) while
1176 \hat{y}_{ij} denotes the scaled simulation output (from training set j). The quantity i denotes
1177 the sampled time-index and \mathcal{T}_j denotes the number of time points for experiment j . In
1178 this study, the experimental data used for model training was typically the band intensity
1179 from Western or Northern blots. Band intensity was estimated using the ImageJ software
1180 package Abramoff *et al.* (2004). The scaled measurement for species x at time $i =$
1181 $\{t_1, t_2, \dots, t_n\}$ in condition j is given by:

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

1182 Under this scaling, the lowest intensity band equaled zero while the highest intensity band
1183 equaled one. A similar scaling was defined for the simulation output. By doing this scal-
1184 ing, we trained the model on the relative change in blot intensity, over conditions or time
1185 (depending upon the experiment). Thus, when using multiple data sets (possibly from

1186 different sources) that were qualitatively similar but quantitatively different e.g., slightly
1187 different blot intensities over time or condition, we captured the underlying trends in the
1188 scaled data. Additionally, we applied a universal convention of identifying the no expres-
1189 sion case as protein value below $<10^{-3}$. This is similar to previously published models
1190 from our lab. In our studies, Figure 2 identifies experimental data extracted from pub-
1191 lished Western blots and our simulation results. It is clear from these that 1) the training
1192 data included a variety of Western blot data treatments that were effectively zero, and 2)
1193 our simulations matched the training data virtually perfectly over time and across multi-
1194 ple biological species. These results validate the power of our simulation scheme, which
1195 necessarily includes numerical interpretations of zero. In our Supplemental Figure S9, we
1196 present the raw concentrations of our simulations (previously validated from the training
1197 data) over hundreds of parameter ensembles, with no “zero” threshold applied. While
1198 these values do not have a threshold, we only interpret values below 10^{-3} equivalent to
1199 zero. Likewise, our robustness coefficients (Figure 3) (which represent the ratio of in-
1200 tegrated areas of the treatment effect over the baseline effect) identify no difference in
1201 model output for vimentin (or e-cadherin for that matter) less than 10^{-3} , confirming this
1202 interpretation.

1203 We computed the Pareto rank of \mathbf{k}_{i+1} by comparing the simulation error at iteration
1204 $i + 1$ against an archive of accepted parameter sets \mathbf{K}_i . We used the Fonseca and
1205 Fleming ranking scheme Fonseca *et al.* (1993) to estimate the number of parameter sets
1206 in the archive that dominate \mathbf{k}_{i+1} . Parameter sets with increasing rank were progressively
1207 further away from the optimal trade-off surface. The parameter set \mathbf{k}_{i+1} was accepted or
1208 rejected by POETs with probability $\mathcal{P}(\mathbf{k}_{i+1})$:

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

1209 where T is the annealing temperature and $rank(\mathbf{k}_{i+1} | \mathbf{K}_i)$ denotes the Pareto rank for
1210 \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	Vimentin	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.

1211 adjusted according to the schedule $T_k = \beta^k T_0$ where β was defined as $(T_f/T_0)^{1/10}$. The
1212 initial temperature was $T_0 = n/\log(2)$, where $n = 4$ in this study and the final temperature
1213 was $T_f = 0.1$. The epoch-counter k was incremented after the addition of 100 members
1214 to the ensemble. Thus, as the ensemble grew, the likelihood of accepting parameter
1215 sets with a large Pareto rank decreased. To generate parameter diversity, we randomly
1216 perturbed each parameter by $\leq \pm 25\%$ at iteration of the search. In addition, we performed
1217 a local pattern search every q -iterations to minimize the residual for a single random or
1218 the worst performing objective function. The local pattern-search algorithm has been
1219 described previously Gadkar *et al.* (2003). From the 15,000 probable EMT models, we
1220 selected $N = 1093$ models with Pareto rank ≤ 1 for subsequent analysis. A quick estimate
1221 of the set to set correlation showed that we could expect on order 25% correlation between
1222 parameter sets in the ensemble.

Mean Trained vs. Cross Validation - (37% significantly different)				Mean Trained vs. Random - (78% significantly different)				Cross Validation vs. Random - (69% significantly different)							
New Feature		S.E. Random		New Feature		S.E. Random		New Feature		S.E. Random		New Feature		S.E. Random	
Mean	0.00	1.00	0.00	Mean	0.00	0.00	0.00	Mean	0.00	0.00	0.00	Mean	0.00	0.00	0.00
Std	0.00	0.00	1.00	Std	0.00	0.00	0.00	Std	0.00	0.00	0.00	Std	0.00	0.00	0.00
Min	0.00	0.75	1.00	Min	0.00	0.00	0.00	Min	0.00	0.00	0.00	Min	0.00	0.00	0.00
Max	-	-	-	Max	-	-	-	Max	-	-	-	Max	-	-	-
CV	-	-	-	CV	-	-	-	CV	-	-	-	CV	-	-	-
CV	0.00	0.00	0.00	CV	0.00	0.00	0.00	CV	0.00	0.00	0.00	CV	0.00	0.00	0.00
VM	0.00	0.00	0.00	VM	0.00	0.00	0.00	VM	0.00	0.00	0.00	VM	0.00	0.00	0.00
Std	-	-	-	Std	-	-	-	Std	-	-	-	Std	-	-	-
BN L2/3	-	-	-	BN L2/3	-	-	-	BN L2/3	-	-	-	BN L2/3	-	-	-

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

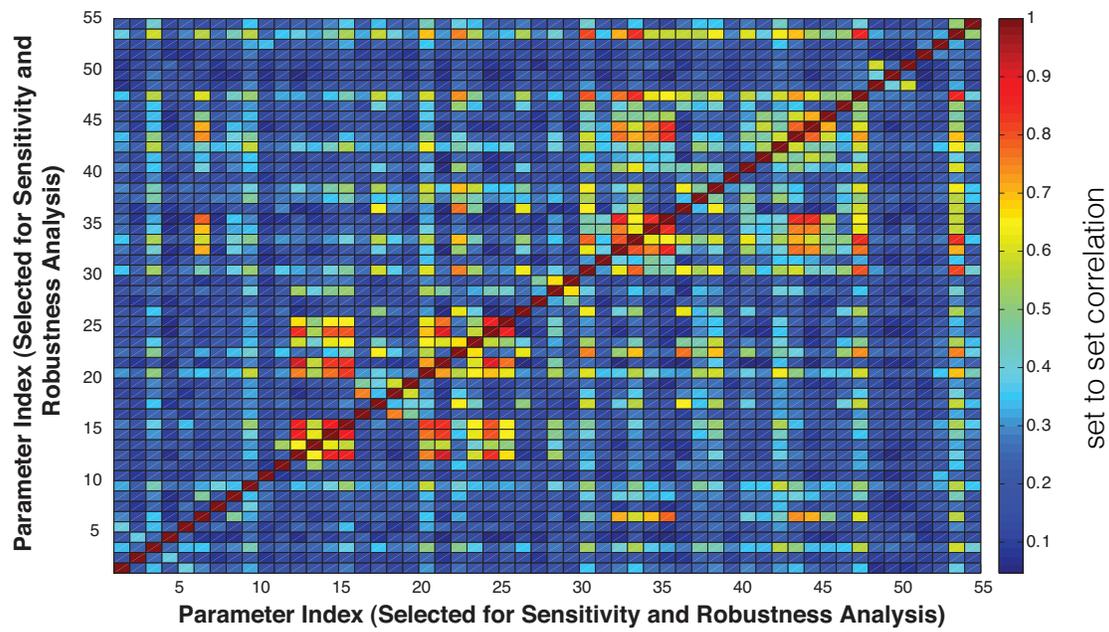


Fig. S7: Parameter set to set correlation for 55 random parameter sets selected from the ensemble. Of the 55 sets selected, the average correlation between sets was less than 25% for greater than 80% of the parameter sets.

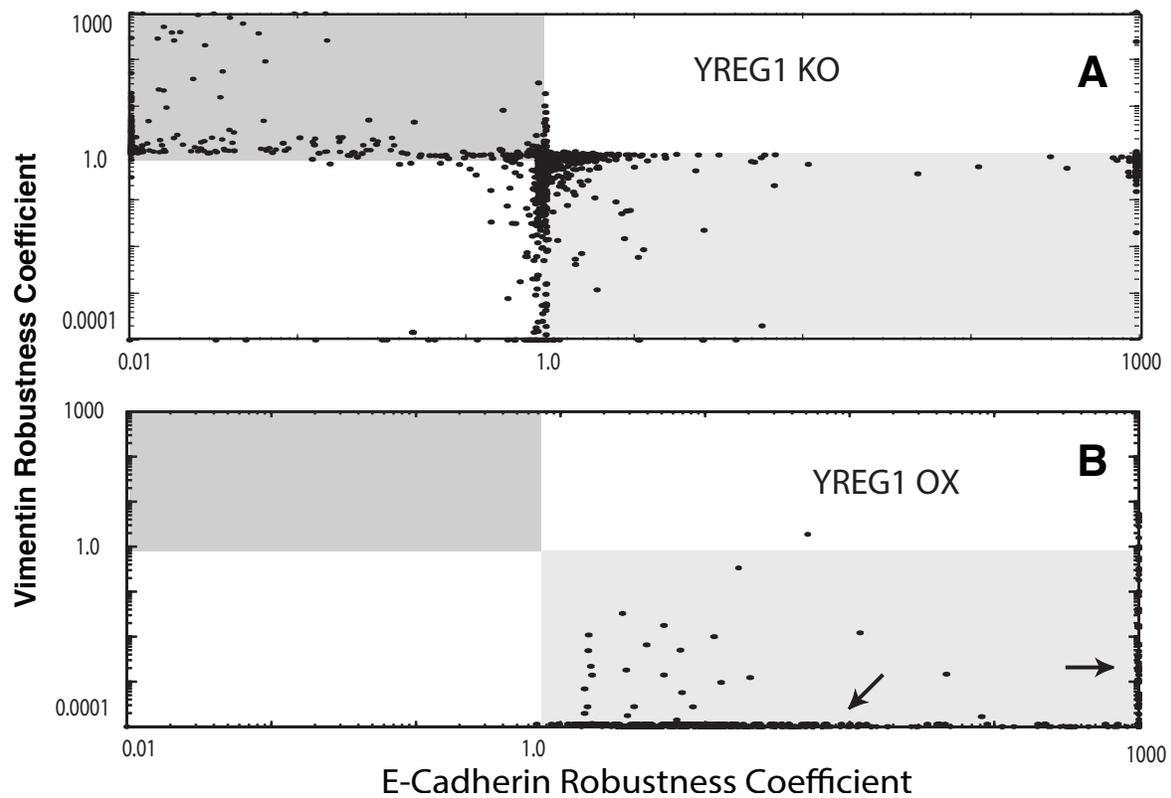


Fig. S8: Robustness of E-cadherin and Vimentin expression to a knockout (A) and overexpression (B) of the hypothetical regulator 1 (YREG1) protein. Robustness coefficients were calculated for each member of the ensemble. Each point represents the response of a single model in the ensemble to either a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

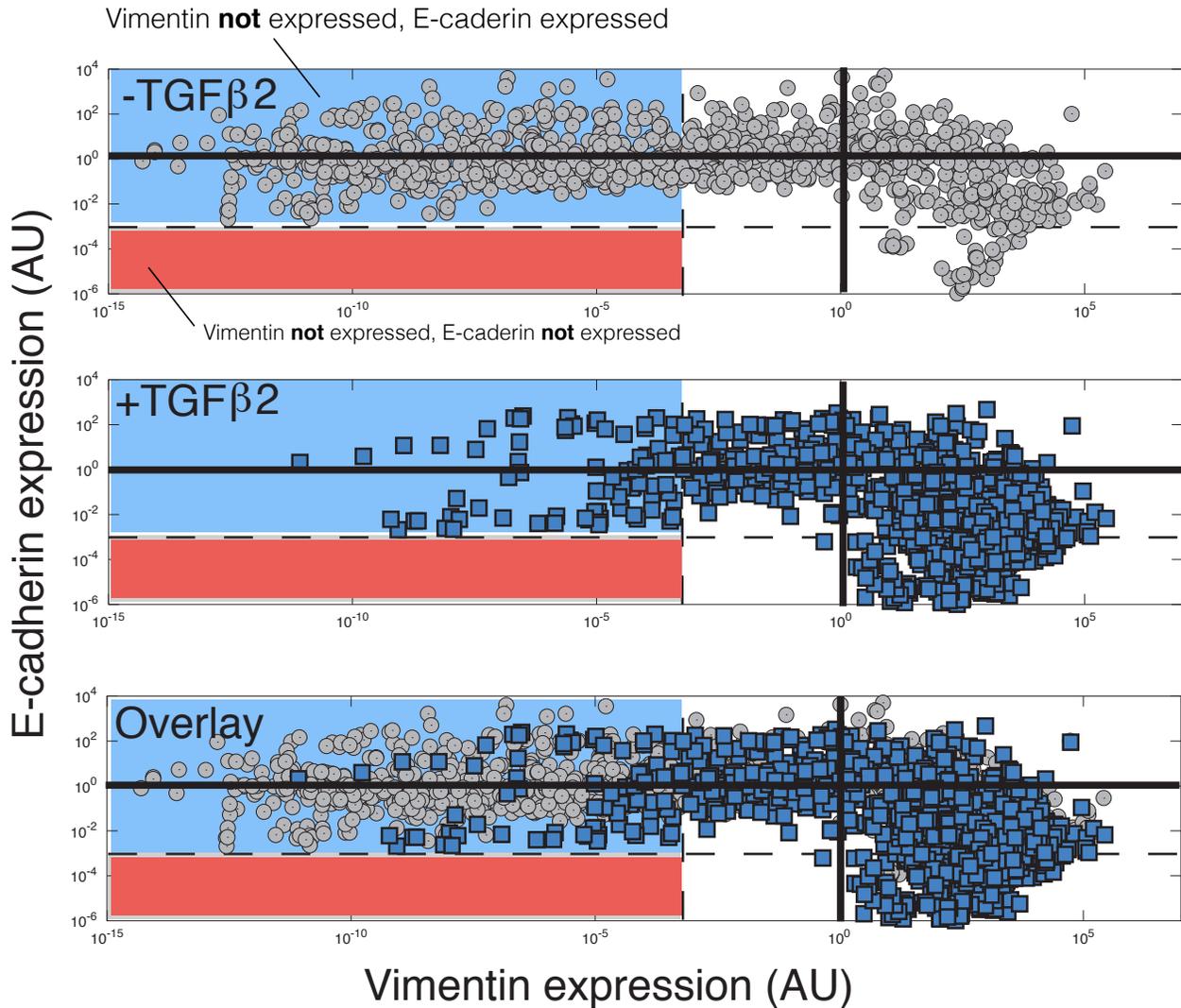


Fig. S9: Steady state protein abundance for E-cadherin and Vimentin (AU) as a function of TGF- β 1/2 exposure. Top: Overlay of the model population for Vimentin (AU) and E-cadherin (AU) expression in the presence (blue) and absence (gray) of TGF- β 1/2. Middle: Vimentin (AU) and E-cadherin (AU) expression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