Network Modeling of TGFβ Signaling in Hepatocellular Carcinoma Epithelial-to-Mesenchymal Transition Reveals Joint Sonic Hedgehog and Wnt Pathway Activation

Steven Nathaniel Steinway, Jorge G.T. Zañudo, Wei Ding, Carl Bart Rountree, David J. Feith, Thomas P. Loughran Jr, and Reka Albert

Abstract

Epithelial-to-mesenchymal transition (EMT) is a developmental process hijacked by cancer cells to leave the primary tumor site, invade surrounding tissue, and establish distant metastases. A hallmark of EMT is the loss of E-cadherin expression, and one major signal for the induction of EMT is TGFβ, which is dysregulated in up to 40% of hepatocellular carcinoma (HCC). We have constructed an EMT network of 70 nodes and 135 edges by integrating the signaling pathways involved in developmental EMT and known dysregulations in invasive HCC. We then used discrete dynamic modeling to understand the dynamics of the EMT network driven by TGFβ. Our network model recapitulates known dysregulations during the induction of EMT and predicts the activation of the Wnt and Sonic hedgehog (SHH) signaling pathways during this process. We show, across multiple murine (P2E and P2M) and human HCC cell lines (Huh7, PLC/PRF/5, HLE, and HLF), that the TGFβ signaling axis is a conserved driver of mesenchymal phenotype HCC and confirm that Wnt and SHH signaling are induced in these cell lines. Furthermore, we identify by network analysis eight regulatory feedback motifs that stabilize the EMT process and show that these motifs involve cross-talk among multiple major pathways. Our model will be useful in identifying potential therapeutic targets for the suppression of EMT, invasion, and metastasis in HCC. Cancer Res; 74(21): 1–15. ©2014 AACR.

Introduction

Hepatocellular carcinoma (HCC) is the third most common cancer type worldwide (1). The only available cure for HCC is surgical resection, which is only an option in early-stage disease. Death from HCC is primarily due to late-stage disease, which is characterized by invasion, intra- and extrahepatic metastasis, and postsurgical recurrence (2). The high mortality and frequency of recurrence in HCC make the need for novel therapeutic approaches paramount. HCC invasion is dependent upon a cellular process called epithelial-to-mesenchymal transition (EMT; ref. 3). Therefore, targeting invasion and the EMT process could reduce HCC-related mortality substantially.

EMT is a crucial developmental event in embryonic tissue formation and regeneration (4). Pathologic forms of EMT are hypothesized to occur during cancer progression and tissue fibrosis (5). During EMT epithelial cells lose their adhesive properties and transform into mesenchymal cells with a migratory phenotype. This provides a critical mechanism in the initial steps of metastasis, allowing cancer cells to leave the primary tumor. Loss of E-cadherin (CDH1), a cell–cell adhesion protein responsible for intercellular attachment, is widely considered the hallmark of EMT (6).

TGFβ (TGFβ1), hereafter referred to as TGFβ signaling has a dual role in tumor progression; it is tumor suppressive in normal epithelia and early-stage tumor cells, contrasted by its stimulation of EMT as cancers progress (7). It is thought that genetic alterations to the cell contribute to the role
Quick Guide to Equations and Assumptions

Major assumptions of the model

The Boolean modeling approach is an abstraction that aims to explain the qualitative characteristics of dynamic systems. The activation of a biologic node is represented as a switch in a Boolean model; the node can be either ON or OFF and there are no intermediate levels of activation. This is a reasonable assumption when a biologic node exhibits two qualitatively different states and the transition between these two states is switch-like.

The regulatory relationships among nodes can be described with the Boolean operators AND, OR, and NOT. If a node has a single regulator then the regulated node’s state is solely determined by its upstream node’s state. OR represents the combined effect of independent and individually effective upstream regulators on a downstream node, whereas AND indicates the joint requirement for multiple upstream regulators to achieve a downstream effect. NOT represents the effect of inhibitory regulators and can be combined with other regulation by using either OR or AND. Our baseline assumption is that multiple regulators of the same target node act independently of each other. However, if there was evidence of biologic synergy (e.g., two proteins form a complex to perform a biologic function), then an AND operator was used. We describe three Boolean equations in Key equations section below; two additional examples are given in the Supplementary Text.

We do not know the kinetics of individual events, although we do know that signal transduction events occur substantially faster (seconds or faster) than transcriptional events (timescale of minutes; ref. 31). Thus, we employ a stochastic asynchronous updating scheme with a ranking system. We randomly select a single node and update its state, meaning that we recalculate its state according to its Boolean equation. To account for signal transduction events occurring faster than transcriptional events, we update nodes regulated by signal transduction events with a greater probability than nodes regulated by transcriptional events. The results are robust to changes in the value of this probability (see Supplementary Fig. S4). By performing a large number (10,000) of simulations with a stochastic update sequence, we identify the behaviors that do not depend on the detailed kinetics of the individual reactions. We define a time step as the average number of updates needed to update a transcriptionally regulated node.

Dynamic Boolean models do not require kinetic parameters, which makes them more computationally feasible than quantitative models, especially for biologic systems with more than a dozen components. Boolean models assume that the structure of the biologic networks they describe is more important than the kinetics of individual reactions and acquire their richness through the large number of interactions that are included in the networks they describe.

Key equations

An explanation of the Boolean regulatory functions of two key nodes is included below. Each equation indicates the future state (shown by the use of the asterisk) of a regulated node as a function of the current states of the nodes that regulate it (i.e., nodes that are the starting points of edges incident on the regulated node). For simplicity, we denote the node state by the node name; thus, “node X” in a rule means “the state of node X,” and “not node X” means “the state that is the opposite of the state of node X.”

We start with a hypothetical example describing the transcriptional regulation of a target gene by a transcriptional activator (A) and a transcriptional inhibitor (I). We denote the transcript of the target gene as T.

\[ T^* = A \text{ and not } I \]  

(A)

The Boolean equation indicates the condition under which the future state of the target gene’s transcript is ON. The right hand side of the equation specifies that this condition is that the transcriptional activator A is ON and simultaneously the transcriptional inhibitor I is OFF. If this condition is not satisfied, then T decays (gets turned OFF), even if it was ON before.

\[ \beta\text{-catenin}_\text{nuc}^* = \text{not Dest}_\text{compl} \text{ and not } \beta\text{-catenin}_\text{memb} \text{ and (not SUFU or not E-cadherin)} \]  

(B)

\[ \beta\text{-catenin}_\text{memb}^* = \text{E-cadherin and not } \beta\text{-catenin}_\text{nuc} \]  

(C)

\( \beta\text{-catenin}_\text{memb} \) is a node representing membrane-bound \( \beta\text{-catenin} \) (CTNNB1). \( \beta\text{-catenin} \) is degraded because of an active “destruction complex” (represented by the Dest\_compl node) formed by APC, AXIN1 or AXIN2, and GSK3\( \beta \). The destruction complex targets free (cytoplasmic and nuclear) \( \beta\text{-catenin} \) for degradation (50); thus, the destruction complex can suppress nuclear \( \beta\text{-catenin} \) (Eq. B; ref. 51). The pool of nuclear (\( \beta\text{-catenin}_\text{nuc} \) node) and membrane-bound \( \beta\text{-catenin} \) (\( \beta\text{-catenin}_\text{memb} \) node) is dynamically controlled by the Wnt signal. Active Wnt signaling leads to \( \beta\text{-catenin} \) localization to the nucleus. Hence, we assume that \( \beta\text{-catenin} \) cannot be simultaneously localized both to the membrane and to the nucleus and implement this assumption as
a mutually inhibitory relationship between β-catenin_memb and β-catenin_nuc (52). We consider an alternative rule for β-catenin_nuc in which there is no inhibition from β-catenin_memb to β-catenin_nuc in Supplementary Fig. S5. Suppressor of fused homolog (SUFU) binds β-catenin, exports it from the nucleus, and thereby represses nuclear β-catenin (Eq. B; ref. 50). E-Cadherin has been shown to suppress β-catenin nuclear localization (Eq. B; ref. 51) and to stabilize membrane-bound β-catenin (Eq. C; ref. 52). We assume that predominant nuclear localization of β-catenin is possible if either SUFU or E-cadherin is absent, if simultaneously the destruction complex is also absent.

switch during tumor progression (8). Previous studies have demonstrated that TGFβ levels are elevated in 40% of patients with HCC (9), whereas receptor levels are variable (10, 11).

TGFβ binds to TGFβRII, which recruits TGFβRI, leading to receptor phosphorylation and phosphorylation of SMAD2 and SMAD3. SMAD2 and SMAD3 form a complex with SMAD4 (in the canonical pathway), which translocates to the nucleus, in which the complex downregulates the transcription of epithelial junction proteins while increasing the expression of mesenchyme-related proteins (6). TGFβ can also activate PI3K/AKT, MAPK, and other established oncogenic signals (noncanonical pathways; ref. 12).

EMT and TGFβ signaling are complex biologic processes that have been extensively studied using reductionist approaches. Network modeling provides a set of tools to study complex processes holistically, using an integration of individual and pairwise evidence (13). A network is a representation of a complex system; the nodes represent entities (proteins, mRNA) or concepts (outcomes) and the edges represent the relationships between nodes. Edges can be assigned a direction, representing information or signal flow from the source (upstream) node to the target (downstream) node. Furthermore, edges can be characterized as positive (activating relationship) or negative (inhibitory relationship). An additional layer of network analysis is the dynamic model, which characterizes each network node with a state variable. The biologic relationships incorporated in the network can be translated into mathematical equations that summarize the regulation of each node. The dynamic model describes how the nodes’ state variables change in time, for example, how a signal activates signal transduction and leads to the transcription of target genes (14).

In this study, we explore TGFβ-driven EMT in HCC by an integrated experimental and computational approach (Fig. 1). We found that TGFβ signaling is a conserved driver of EMT across multiple HCC models. We developed a network model of EMT signaling and then translated this network into a discrete dynamic model. The dynamic model recapitulates the known steps of TGFβ-driven EMT. The model predicts, and our experiments confirm, that joint activation of Wnt and Sonic hedgehog (SHH) signaling by TGFβ is a conserved feature of TGFβ signaling in EMT. Our analysis suggests a putative reliance on these pathways for robust activation of the EMT program.

Materials and Methods

Cell culture

P2E and P2M cells were cultured in DMEM:F12 1:1 medium (Mediatech) supplemented with 10% FBS as described previously (15). Huh7 cells were provided by Dr. Harriet Isom (Penn State College of Medicine, Hershey, PA) and were cultured in DMEM + 10% FBS medium (16). Human HLE cells were acquired from Dr. Curtis Harris (National Cancer Institute) and cultured in DMEM + 10% FBS. HLF cells were acquired from Dr. Jorge Filmus (University of Toronto, Toronto, ON, Canada) and cultured in DMEM + 10% FBS (17). PLC/PRF/5 cells were acquired from the ATCC and cultured in Eagle’s Minimum Essential Medium + 10% FBS. TGFβ inhibitor treatments were completed with LY2157299 (Selleck Chemical).

Cell line authentication

Huh7, HLE, and HLF cell lines were authenticated using short tandem repeat DNA profiling (Genetica DNA laboratories) on January 20, 2014. PLC/PRF/5 cells were received directly from the ATCC (July, 2013) and cultured for under 3 months so no authentication was deemed necessary.

Quantitative real-time PCR

Real-time PCR experiments were conducted as described previously (18). The housekeeping gene GAPDH was used for ΔΔCt calculations and relative expression was calculated using TaqMan primer/probe sets (Life Technologies).

Western blot analysis

Lysates were harvested with lysis buffer as described previously (18). Of note, 50 μg of protein lysates was separated on a NuPAGE 4% to 12% Bis-Tris Gel (Invitrogen) and transferred to polyvinylidene difluoride membrane (Invitrogen). Primary antibodies for E-cadherin (Cell Signaling Technology), pSMAD2 (Cell Signaling Technology), pSMAD3 (Cell Signaling Technology), SMAD2 (Cell Signaling Technology), SMAD3 (Cell Signaling Technology), β-actin (Cell Signaling Technology), GLI2 (Santa Cruz Biotechnology), and rabbit and mouse secondary antibodies (Cell Signaling Technology) were used for this study. Signals were detected using Clarity chemiluminescence substrate on a ChemiDoc XRS+ imaging system (Bio-Rad).

Microarray gene set enrichment analysis

Murine HCC P2E and P2M cell line microarray datasets were acquired from NCBI Gene Expression Omnibus (15). Gene set enrichment analysis (GSEA) was performed to determine gene sets representing pathways and phenotypes enriched in P2M relative to P2E cell lines using datasets in the Molecular Signatures Database 4.0 (19). Further details of GSEA are described in the Supplementary Text.
Patient samples

Human HCC samples were acquired from liver biopsy at Penn State Hershey Medical Center (Hershey, PA). Patients were on no cancer treatment at acquisition and were diagnosed as primary HCC. Specimens were flash-frozen, and protein was extracted by homogenization of 100-mg liver tissue in 1-mL extraction buffer (150 mmol/L NaCl, 50 mmol/L Tris/HCl pH 7.4, 1 mmol/L EDTA, 1% Triton 100, 0.1% SDS, 1% sodium deoxycholate, 1 mmol/L sodium pyrophosphate, 10 mmol/L β-glycerol phosphate, 10 mmol/L NaF, 0.5 mmol/L NaVO4, 1 μmol/L Microcysteine-LR, and Protease inhibitor Cocktail Tablet without EDTA (1 tablet in 10 mL lysis buffer)). Samples were incubated on ice for 30 minutes and vortexed every 5 minutes. Samples were centrifuged at 15,000 rpm at 4°C for 15 minutes. Protein (supernatant) was collected and quantified using BCA Protein Assay (Thermo Scientific).

Computational methods

An overview of the computational methodology is included in Fig. 1 and in the Quick Guide to Equations and Assumptions. Network simplification was performed using analytic procedures (20) and with NET-SYNTHESIS, a signal transduction network inference and simplification tool (21). The simulations of the model and the state space analysis were performed using custom Java code and Python code (BooleanNet Python library; ref. 22). The feedback loops responsible for the steady state of the EMT network were identified by stable motif analysis. This method is based on the principle that a certain group of nodes that forms a network motif with identifiable properties stabilizes into a single state regardless of the rest of the network. The specific criteria for identifying stable motifs were described previously (23), and adaptation to our network is described in the Supplementary Text. The stable motif analysis was implemented in a custom Java code. The networks in Fig. 3, Fig. 5A and C, Fig. 7A and C, and Supplementary Fig. S3 were created using the yEd graph editor by yWorks (http://www.yworks.com/). The state transition network in Fig. 5B was generated with Cytoscape (24) by using the Prefuse force-directed layout.

Figure 1. Flowchart of computational algorithms. Our strategy was the following: we knew that TGFβ is a prominent signal driving EMT in HCC, so we started with synthesizing a network for this process from the literature. Then, we added other known inducers of EMT and processes and components that are known to be deregulated during EMT. The resulting EMT network contains 70 nodes and 135 edges (EMT network; Fig. 3). We formulated a discrete dynamic model of this network by using our best estimation of the Boolean functions describing the regulation of each node. We then used simulations of the model to reveal the possible outcomes of TGFβ signaling in a cell that initially exhibits an epithelial phenotype (dynamic simulations; Fig. 4). To explore the full range of possibilities in terms of outcomes (steady states) the network can reach and ways in which these outcomes can be reached (trajectories), we performed a state space analysis. This analysis had to be preceded by network reduction to reduce the size of the network without disrupting its dynamic repertoire. State space analysis on the reduced network (reduced the EMT dynamic model; Fig. 5) confirmed the results of the simulations. Furthermore, we used stable motif analysis to identify key feedback loops regulating EMT (motif analysis; Fig. 7A).
Figure 2. Active TGFβ signaling correlates with mesenchymal phenotype in murine and human models of HCC. A, treatment of a mouse epithelial HCC cell line (P2E) with TGFβ1 at 5 ng/mL for 48 hours produces a molecular phenotype similar to the mouse mesenchymal HCC cell line (P2M). B, upregulation of mRNA levels of EMT markers in mouse epithelial HCC with TGFβ1 treatment (5 ng/mL for 48 hours). C, GSEA supports that P2M cells are enriched for mesenchymal markers relative to P2E cells (left). GSEA reveals that TGFβ signaling is enriched in P2M versus P2E cells, suggesting that TGFβ may be driving the EMT phenotype in this mouse model of HCC (right). This gene set contains genes upregulated by TGFβ1 in MCF10A cells (false discovery rate = 0.031). D, human mesenchymal HCC cell lines HLE and HLF have low E-cadherin expression and high pSMAD3, whereas epithelial-like HCC cell lines Huh7 and PLC/PRF/5 (PLC) express high E-cadherin and low pSMAD3. E, relative to the PLC/PRF/5 cell line, PLC/PRF/5 treated with TGFβ (10 ng/mL, 48 hours), HLE, and HLF cell lines exhibited lower E-cadherin and higher vimentin, and TWIST1 expression.
Results

**TGFβ signaling is a conserved driver of EMT in multiple liver cancer models**

Our initial studies involved development of a novel murine model of EMT. This involved derivation of "sister" epithelial (P2E) and mesenchymal/metastatic (P2M) cell lines from a *PTEN*<sup>−/−</sup> sideline (15). Consistent with other carcinoma cell lines, treatment of the P2E cell line with TGFβ induced TGFβ signaling (Fig. 2A) and led to induction of EMT markers (Fig. 2B). These results demonstrate that activation of TGFβ signaling recapitulates the mesenchymal phenotype in a murine model of HCC. They also led us to hypothesize that the TGFβ pathway is constitutively active in P2M cells as compared with P2E cells.

We tested this hypothesis by GSEA, a computational method that determines whether an *a priori* defined set of genes shows statistically significant, concordant differences between two phenotypes, in this case P2E and P2M cell lines (19). The primary result of GSEA is the enrichment score, which shows the degree to which a gene set is overrepresented at the top or bottom of a list of genes ranked by their association with a phenotype (see Materials and Methods). An EMT gene set representing genes upregulated during EMT induced in a mammary epithelium cell line transformed by *HRAS* (EpH4...
line; ref. 25) was used to probe for the EMT phenotype in P2M versus P2E cells. The use of a breast cancer reference set is reasonable because EMT is a conserved process over multiple tissue types (26). A TGFβ gene set representing genes upregulated by TGFβ1 in MCF10A immortalized nontumorigenic mammary gland cells (27) was used to probe for TGFβ pathway activation in P2M versus P2E cell lines. The top portion of the plot shows the running enrichment score and a maximum enrichment score of 0.78 for the EMT gene set and 0.76 for the TGFβ gene set (Fig. 2C). In the bottom portion of the plot, P2M cells are represented on the left of the figure and P2E on the right. The position of the members of the gene set in the ranked list of genes demonstrates distinct genes associated with EMT and TGFβ sets. These data support upregulation of EMT and TGFβ pathway activation gene sets in the P2M cell line relative to P2E (Fig. 2C).

To explore the prevalence of dysregulated TGFβ signaling in HCC, we tested established human HCC cell lines for TGFβ pathway and EMT markers. Human HCC cell lines Huh7, PLC/PRF/5, HLE, and HLF cell lines were tested for epithelial and mesenchymal markers to characterize their phenotype. Huh7 and PLC/PRF/5 cells have higher E-cadherin and lower pSMAD3 compared with HLE and HLF cell lines by Western blot analysis (Fig. 2D). TGFβ-treated PLC/PRF/5 cells, HLE and HLF HCC cell lines have lower E-cadherin, higher vimentin (VIM), and higher TWIST1 compared with PLC/PRF/5 cells as indicated by quantitative real-time PCR (qRT-PCR; Fig. 2E). Compared to PLC/PRF/5 cells, TGFβ treated PLC/PRF/5 cells and untreated HLF cells have higher N-cadherin expression (Fig. 2E). These results indicate that the Huh7 and PLC/PRF/5 cell lines have an epithelial-like phenotype, the HLE and HLF cell lines have a mesenchymal phenotype.

Figure 4. Network modeling reproduces established signaling abnormalities seen in TGFβ-driven EMT and predicts novel cross-talk mechanisms. A, heatmap representing the activity (averaged state) of all nodes according to the model (10,000 simulations, asynchronous update). The initial state corresponds to a prototypical epithelial state; the final state indicates a TGFβ-driven EMT steady state. Nodes whose state did not change during the simulation are not included. Node activity level of 0 (black square) indicates that the node is inactive (OFF) in all the simulations. Node activity level of 1 (white square) indicates that the node is active (ON) in all the simulations. B, time-lapse dynamic simulation (10,000 simulations, asynchronous update) reveals induction of SNAI1 expression and loss of E-cadherin expression during induction of EMT from an epithelial initial condition. C, time-lapse dynamic simulations (10,000 simulations, asynchronous update) reveal induction of SHH signaling and downstream GLI expression, Wnt signaling, and AXIN2 expression with induction of EMT from an epithelial initial condition.
phenotype, and that TGFβ treatment induces a mesenchymal-like phenotype.

To study the effect of TGFβ pathway activity on the EMT phenotype of the mesenchymal-like HLE and HLF cell lines, these cell lines were treated with TGFβRI inhibitor LY2157299 and EMT markers were assessed by Western blot analysis and qRT-PCR. TGFβR1 inhibition with LY2157299 reduced SNAI1, TWIST1, N-cadherin (CDH2), and vimentin mRNA expression, while increasing E-cadherin expression in HLE cells (Supplementary Fig. S1). These data support the conclusion that the mesenchymal phenotype of HLE and HLF cell lines is due to constitutive TGFβ signaling.

Overall, our data demonstrate that TGFβ axis dysregulation is a conserved driver of mesenchymal phenotype in multiple models of HCC.

Constructing the signaling network that drives EMT

To explore the systems-level dynamics of TGFβ signaling and EMT, a network modeling approach was used. Evidence supports the concept that a conserved EMT mechanism, driven by multiple signals, occurs in development and in cancer. These signals appear as a normal part of development, playing a crucial role in formation of the body plan and in differentiation of multiple tissues and organs (6). During carcinoma progression, EMT signals (e.g., TGFβ) are aberrantly produced and hijack this developmental process (28). To construct our network, we performed an extensive literature search and synthesized preexisting information representing the conserved regulation of EMT. We focused on interactions from HCC EMT; however, information about EMT in other tissue types was also used when no HCC-specific information
was available. The resulting network (Fig. 3) spans heterogeneous molecular processes induced by growth factors, signal transduction pathways, and transcriptional regulators. This network includes the dysregulation of the TGFβ/TGFβR axis as a driver of EMT, as this is a commonly mutated pathway in HCC (9).

The information used to construct the network is summarized in Supplementary Tables S1 and S2. Proteins, mRNAs, and microRNAs were represented as nodes (Supplementary Table S1) and the pairwise relationships among nodes (Supplementary Table S2) were represented as edges. We choose as the output node of the network the concept node "EMT," serving as the major indicator of cell fate in our model. E-cadherin plays a crucial and direct role in EMT at diverse developmental stages and during carcinoma progression, and its loss is widely considered as the hallmark of EMT (6). Furthermore, the regulation of E-cadherin expression is much better understood than that of other EMT markers (29). Thus, we focus on E-cadherin as the negative regulator of the output node of the network, "EMT." This network contains 70 nodes and 135 edges (Fig. 3).

Translating the EMT signaling network into a dynamic model

To understand the dynamics of signaling abnormalities in HCC progression, invasion, and metastasis, we translated the EMT signaling network into a dynamic model. A dynamic model is used to express the behavior of a system over time by characterizing each node by a state variable (representing, e.g., its concentration or activity) and a function that describes its regulation. Dynamic models can be categorized as continuous or discrete, according to the type of node state variable used. Continuous models use a set of differential equations; however, the paucity of known kinetic details for biologic interactions makes these models difficult to implement.

We used discrete dynamic modeling because of its computational feasibility and capacity to be constructed with qualitative biologic data. Specifically, we constructed a Boolean model wherein each network node is described by one of two qualitative states: ON or OFF. The ON state means an above threshold concentration of a molecular regulator whereas the OFF state means the below threshold or inactivated form. We formulated a regulatory function for each node using its upstream regulators encapsulated in the network, evidence regarding the conditionality among these regulators, and known biologic outcomes. During the construction of the regulatory functions, we verified that individual modules within the network reproduced known outcomes (Supplementary Table S4). If modules did not reproduce known outcomes, we modified the rules in such a way that the known outcomes can be reproduced (30). See the Quick Guide to Equations and Assumptions and Supplementary Text for more detailed information and example rule implementations.

As in the biologic system, there is a time lag between the state change of the regulators and the state change of the targets. This time lag is implemented by using discrete update events, each update meaning that the state of a node is recalculated according to its Boolean equation. We used an updating scheme that differentiates between the two major types of biologic regulation in the model: transcriptional events and signal transduction events. Although we do not know the specific kinetics of individual events, we do know that signal transduction events occur substantially faster (in seconds or faster) than transcriptional events (timescale of minutes; ref. 31). Thus, we used a ranked asynchronous updating algorithm that updates nodes regulated by signal transduction events with a greater probability than nodes regulated by transcriptional events. This allows us to stochastically explore differences in the kinetics among molecular interactions in our network while satisfying known kinetic relationships (14). The details of the ranked asynchronous algorithm are described in the Supplementary Text.

To reproduce how a potentially diverse population of cells responds to the same signal, we performed multiple simulations with the same initial conditions but different updating orders (i.e., different timing). The output of the model is the fraction of simulations in which a certain node is in the ON state, in which we denote the activity of the node. We chose as a time unit (step) the average number of updates needed to update a transcriptionally regulated node. This time step, thus, corresponded to several minutes in real-time. We empirically determined that 10,000 simulations and 20 time steps in each simulation were sufficient for convergent and reproducible results (see Supplementary Text for further explanation).

The EMT network model reproduces known TGFβ-driven dysregulations during EMT

We focus on TGFβ as the prototypical inducer of EMT as the TGFβ signaling axis is disrupted in 40% of HCC (9) and TGFβ is a well-established inducer of EMT in HCC and other carcinomas (8). First, we set the states of the nodes according to their known or expected states in a prototypical epithelial cell. Specifically, known signals/pathways that could induce EMT, including TGFβ, FGF, IGF, EGF, PDGF, Wnt/β-catenin, SHH/GLI, and HGF/cMet, were set to OFF. Downstream signal transduction pathways known to drive EMT, such as MAPK, AKT, SMAD, and STAT signaling, were also turned OFF. E-Cadherin and miR200 were assumed to be ON and the E-cadherin transcriptional repressors were assumed to be OFF. Our model indicates that without any stimulus or dysregulation, the cell would remain in this epithelial state (listed in Supplementary Table S3).

Next, to recapitulate a sustained TGFβ signal, the state of TGFβ was fixed in the ON state at the beginning of every simulation (see "Boolean rule over-rides" in the Supplementary Text), whereas the other nodes were in their epithelial initial condition. Turning TGFβ ON led to changes in the activity of nodes in the network over the time course of the simulation and ultimately led to a steady state in which the state of all nodes stabilized (Fig. 4A). If the state of a node stabilized at ON at the end of the simulation, even though it was OFF at the beginning, we infer that it is activated in the steady state. If the state of a node was OFF even though it was in the ON state at the beginning of the simulation, we consider it inactivated in the steady state. Using this interpretation, the obtained steady state agrees with known markers of a mesenchymal state. The
heatmap in Fig. 4A and the time courses shown on Fig. 4B qualitatively reflect experimental outcomes seen when starting from an epithelial condition and inducing EMT. Specifically, they reflect how TGFβ signals through TGFβRI/II and induces SMAD complex formation and noncanonical pathways, including MAPK and AKT. This leads to induction of E-cadherin transcriptional repressors and ultimately to the loss of E-cadherin (Fig. 4A and B). The agreement of our network model with known experimental outcomes of TGFβ-driven EMT supports the validity of our model.

To explore the possibility of other steady states besides the previously described EMT steady state, we performed a state space analysis of the model. The state space can be represented as a network (the state transition network), whose nodes represent the system’s states and whose edges represent every state transition that is possible for the given system. The state transition network summarizes all the possible trajectories or ways that a network state can lead to another network state. Steady states of the system are states that only have transitions leading to them but not leading out of them. The state space grows exponentially with the size of the network, making it impossible to fully map for large networks. To decrease the state space, we performed network simplification of the 70 node, 135 edge EMT network in such a way that the network remains consistent with causal experimental observations (See Supplementary Text for details; ref. 32). This network reduction has been shown to have no effect on the permitted dynamic behaviors of a system (33). This led to a network with 19 nodes and 70 edges (Fig. 5A and Supplementary Table S5).

In the reduced 19 node network, TGFβ was fixed as ON, and the part of the state space reachable from the otherwise epithelial initial condition was determined, using a ranked asynchronous updating algorithm. The resulting partial state transition network has 3,489 nodes (states) and 16,434 edges (state transitions; Fig. 5B). Only one node of the state transition network has no outgoing edges, the node corresponding to the EMT steady state. This, steady state is the only possible outcome for an epithelial cell in the presence of sustained TGFβ signaling; no other steady states were discovered. This steady state can be reached through many thousands of trajectories, each corresponding to a different set of state transitions. For example, the four most common trajectories (Fig. 5C) include E-cadherin loss before nuclear localization of β-catenin, or following it. As trajectories in the model reflect the order of events possible under different kinetics, their convergence into a single outcome suggests that the TGFβ-driven EMT outcome is robust across all reasonable values of kinetic parameters. We note that none of the scenarios to reach EMT encapsulated in the state transition network involve specific dysregulations (e.g., mutational gain or loss of function) of any nodes in the EMT network. The presence of constitutive TGFβ signaling drives all dysregulations we see in our model.

Network modeling reveals that the constitutive presence of TGFβ induces Wnt and SHH signaling during EMT in HCC

Interestingly, our dynamic model of signaling induced by TGFβ activation revealed joint activation of Wnt and SHH signaling pathways during induction of EMT from an epithelial initial condition (Fig. 4A and C). To test the predictions from our model, we explored molecular readouts of Wnt and SHH signaling. AXIN2 is a well-established transcriptional marker of Wnt/β-catenin/TCF signaling and is a negative regulator of the Wnt signaling pathway (34). Furthermore, members of the GLI family of transcription factors (represented by the node GLI in our model) are major effectors of the SHH signaling pathway, and their mRNA expression can be used as a readout of SHH signaling (35). GLI2 is the major transcriptional activator of the SHH signaling pathway, with GLI2 itself undergoing transcriptional upregulation when SHH signaling is active (36). Most evidence supports that GLI1 is activated transcriptionally by GLI2 (37). Indeed in our model, simulations (10,000 simulations, asynchronous update) starting from an epithelial state in the sustained presence (ON state) of Wnt led to AXIN2 expression (Supplementary Fig. S2A) and the presence of SHH in the same initial state drove GLI expression (Supplementary Fig. S2B). These results affirm that our model reproduces known molecular readouts of Wnt and SHH signaling.

As shown in Fig. 4C, TGFβ signaling induces both AXIN2 and GLI (with a delay compared to SHH) in our model. To experimentally explore the activation of Wnt and SHH signaling pathways by TGFβ, we tested TGFβ-driven expression of pathway markers using epithelial and mesenchymal cell lines. Treatment of the P2E cell line (murine epithelial HCC cell line) with TGFβ leads to induction of GLI2 mRNA and protein expression and induction of AXIN2, approximating the levels constitutively observed in mesenchymal P2M cells (Fig. 6A and B). TGFβ treatment drives GLI1 and GLI2 expression in human epithelial-like HCC cell lines Huh7 and PLC/PRF/5 cells. Furthermore, AXIN2 is upregulated by TGFβ in Huh7 and PLC/PRF/5 cells after TGFβ treatment. In addition, human mesenchymal HCC cell lines HLE and HLF exhibit upregulated GLI expression relative to epithelial PLC/PRF/5 cells (Fig. 6C). Inhibition of TGFβ signaling with a chemical inhibitor of TGFβRI/II diminished pSMAD3 and GLI2 protein expression in the mesenchymal HLE and HLF cell lines (Fig. 6D). Thus, our measurements validate the model prediction that Wnt and SHH signaling is induced by TGFβ during induction of EMT.

To further test the joint activation of SHH signaling pathways by TGFβ, we determined the correlation between TGFβ and GLI2 expression in human HCC patient samples. Indeed, SMAD3 phosphorylation correlated positively ($R^2 = 0.629$) with GLI2 expression in these samples (Fig. 6E). This information supports the network model’s prediction that induction of GLI2 occurs in human tumors with active TGFβ signaling.

For a deeper analysis of the nodes associated with steady states in the EMT network, a stable motif analysis was performed. This analysis is based on finding feedback loops that stabilize in a fixed state (23). Our findings further demonstrate the existence of a single EMT steady state (i.e., the state shown on Fig. 4A). There are eight motifs associated with this steady state (Fig. 7A and Supplementary Fig. S3). These motifs represent interaction and node activity patterns whose sustained presence is sufficient to stabilize the mesenchymal state. The motif shown on the left of Fig. 7A indicates that sustained Wnt and GLI signaling and a low level of GSK3β can maintain a low
level of E-cadherin, the hallmark of the EMT steady state. The motif in the middle of Fig. 7A indicates that a low level of GSK3β can be sustained by cross-talk between MAPK and SMAD signaling. The motif on the right side of Fig. 7A indicates that SMAD signaling can be activated by the TGFβ signal in our model, which itself can be sustained by positive feedback. Read from right to left, these motifs are consistent with the propagation of the TGFβ signal that leads to SMAD and SNAI1 activation, then to the activation of the SHH and Wnt pathways, and then to the loss of E-cadherin. However, because each stable motif is independent of the others and is sufficient to maintain the EMT steady state, the results shown on Fig. 7A suggest that other, potentially TGFβ independent, mechanisms that lead to the stabilization of a motif (e.g., MAPK signaling) can also drive EMT. Indeed, evidence for the induction of EMT through MAPK pathway activation of SNAI1 has been reported in other studies (38). In our model, MAPK signaling can drive EMT, independent of TGFβ (data not shown). Taken together, our studies demonstrate that activation of SHH and Wnt pathways are a key conserved feature of HCC EMT.

Discussion

We began our studies when we realized that TGFβ signaling mimics the molecular phenotype of a murine mesenchymal cell counterpart (P2M) to an epithelial HCC cell line (P2E). We next identified other HCC cell lines with active TGFβ signaling and demonstrated that TGFβ-active cell lines also had mesenchymal features. These results suggest that TGFβ signaling is
a conserved dysregulation and driver of EMT across multiple HCC models. To better understand the complex signaling events producing the TGFβ-driven EMT phenotype, we curated signaling pathways involved in developmental and carcinoma EMT programs in HCC and compiled them into an EMT signaling network. By formulating a Boolean dynamic model of this network, we were able to identify downstream signaling events involved in the TGFβ-specific EMT process and detect cross-talk among pathways. As EMT and TGFβ signaling are conserved features of many solid tumor types, this model (or modifications of it) may be used to study EMT in other tumor types.

TGFβ plays a dual role as a tumor suppressor and oncogene; however, it remains unclear why. We argue that in the cell, network-level emergent outcomes (e.g., tumor suppression vs. tumor growth) are not attributable to the TGFβ signal alone, but are dependent on the status of the entire cellular network that the signal acts upon. Network models are increasingly used to understand the organization and emergent properties of biologic systems (39, 40). Network-based discrete dynamic models are particularly useful in poorly characterized biologic systems, in which they can be used to generate testable hypotheses (41). Boolean network models have led to new insight into signal transduction and gene regulatory networks in numerous organisms (14). Network modeling can explain why in certain differential contexts (e.g., for different states of a few nodes other than TGFβ), the TGFβ signal produces different outcomes, which can then be investigated experimentally.

Perturbations in signaling pathways that define a disease or a specific feature of a disease can result from deregulation of only a subset of these pathways. However, identifying such a subset is difficult using traditional biologic approaches. In this study, we assessed this question in EMT, a mechanism that drives invasion and metastasis in solid tumors, by using a Boolean dynamic model. Through network simulations, we revealed that the TGFβ signal could percolate across multiple signaling pathways, leading to a network of perturbed proteins
during the induction of EMT. Across multiple murine and human HCC models, the constitutive presence of TGFβ was shown to be sufficient to induce Wnt and SHH signaling pathways in this context. The model’s result regarding the order of induction of the Wnt and SHH pathways obtained from the model may depend on the broad classification used for the time lags (transcriptional vs. signal transduction events). A future experimental determination of the order of pathway activation may support the model result or may contradict it, in which case, further timing constraints could be incorporated to establish agreement. Imposing timing constraints would affect how fast the TGFβ signal propagates through the network but it would not affect the network’s steady states.

It should be noted that TGFβ, Wnt, and SHH signaling pathways have individually been implicated in oncogenic processes, including EMT in carcinomas. For example, Wnt pathway induction of EMT has been demonstrated in multiple cancer types (42). Furthermore, SHH can induce EMT in brain and gastrointestinal cancer models, among others (43). It has recently been suggested that the critical interaction between TGFβ and Wnt signaling pathways deserves special attention (44). Here, we present a systems level explanation of the interaction of TGFβ signaling with the Wnt and SHH signaling pathways. How these signaling pathways act as a network to integrate a signal for a specific biologic process has not been realized by previous studies. Our stable motif analysis revealed that the SHH/GLI and Wnt components of the EMT network are critical nodes in feedback loops that maintain the EMT steady state (Fig. 7A).

EMT is a process that occurs throughout development and is hijacked by carcinomas during solid tumor progression. Tumor metastasis is the leading cause of mortality in solid tumors. Our model provides a strong foundation for the modeling of HCC invasion. As hepatic tissue lacks a membrane barrier, it has been speculated that EMT is the major mechanism for the entry of migratory HCC cells into systemic circulation (3). To describe carcinomas of epithelia in other tissues, which do have a basement membrane, our model can be expanded by adding the pathway(s) regulating the proteolytic destruction of the basement membrane by matrix metalloproteinase proteins (45). Another possible extension is an agent-based model in which each cell is an agent with a specific morphology. Such a model could incorporate our Boolean signal transduction model, the initial epithelial architecture, as well as the reorganization of the actin cytoskeleton of each cell to enable directional motility and dynamic cell elongation (46).

Understanding the systems-level dysregulations in EMT is important not only to elucidate tumor progression and the metastatic process but also to develop therapeutic interventions to halt/suppress metastatic disease. Our analysis suggests that single interventions may not be effective. For example, SMAD proteins are the canonical downstream targets of TGFβ signaling. However, our model predicts that inhibiting SMAD proteins alone cannot completely block TGFβ-driven EMT (Fig. 7B). Indeed, among the motifs that can sustain an EMT = ON state there is a TGFβ feedback loop that does not include SMAD (Supplementary Fig. S3). Stable motif analysis with TGFβ = ON and SMAD = OFF (a TGFβ active network with no SMAD signaling) reveals multiple feedback loops that can maintain EMT in the absence of SMAD signaling (Fig. 7C). This implies that combining SMAD inhibition with Wnt, SHH, AKT, or MAPK inhibition may more effectively suppress EMT. Similar computational analyses have been completed on smaller networks, supporting a role for feedback loops in the regulation of EMT (29), EMT plasticity (47), and tumor cell heterogeneity (nongenetic heritability; ref. 48).

Combinatorial therapeutics is an emerging discipline in the treatment of patients with cancer. Recent work indicates that targeting multiple oncogenic signals through combination of cancer therapies is superior to single-therapy treatment (49). The development of effective combined therapeutics is an important goal within this field. Developing dynamic models of cancer networks to simulate single and combinatorial interventions targeting specific nodes may prove to be a rational approach to more efficiently choose single and multidrug therapeutic regimens.

In summary, we used network analysis and Boolean modeling to investigate the signaling abnormalities that arise during EMT in the context of TGFβ signaling and confirmed our network-based results in multiple experimental liver cancer models (Fig. 1). Our systems biology approach was able to maximize the use of the available pathway information and to identify putative activation of major signaling pathways during EMT and liver cancer invasion. This study confirms the ability to integrate normal and disease pathway information into a single model that is robust enough to reproduce a clinically relevant complex process yet is constructed with qualitative information.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Disclaimer
The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors’ Contributions
Conception and design: S.N. Steinway, C.B. Rountree, D.J. Feith, T.P. Loughran Jr, R. Albert
Development of methodology: S.N. Steinway, J.G.T. Zaidulo, C.B. Rountree, R. Albert
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.N. Steinway, W. Ding
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.N. Steinway, J.G.T. Zaidulo, T.P. Loughran Jr, R. Albert
Writing, review, and/or revision of the manuscript: S.N. Steinway, J.G.T. Zaidulo, W. Ding, C.B. Rountree, D.J. Feith, T.P. Loughran Jr, R. Albert
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.N. Steinway, C.B. Rountree, T.P. Loughran Jr
Study supervision: C.B. Rountree, D.J. Feith, T.P. Loughran Jr, R. Albert

Acknowledgments
The authors thank Dr. Hien Dang (NIH/NCI) for training and technical expertise at the onset/during pilot studies. The authors also thank Dr. Xin Liu (Penn State College of Medicine) for antibodies.
Grant Support
This work was supported by NIH F30DK093234 (S.N. Steinway), NSF IIS 1161001 (R. Albert), and NSF PHY 1205840 (R. Albert).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 24, 2014; revised June 27, 2014; accepted August 18, 2014; published OnlineFirst September 4, 2014.

References
Network Modeling of TGFβ Signaling in Hepatocellular Carcinoma Epithelial-to-Mesenchymal Transition Reveals Joint Sonic Hedgehog and Wnt Pathway Activation

Steven Nathaniel Steinway, Jorge G.T. Zañudo, Wei Ding, et al.

Cancer Res  Published OnlineFirst September 4, 2014.

Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-14-0225

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/09/05/0008-5472.CAN-14-0225.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.