We thank the reviewers for their comments and are excited for their enthusiasm for our submission. We have considered both their comments and those of the editor, and have revised our submission. In particular, we have revised figures and added clarifying text. Our point by point response to each comment is below in blue font. Our corresponding text changes in the revised manuscript are highlighted in yellow.

**Editor’s Comments**

*“From the authors' response to my previous question about the size of the robustness coefficients, I remain confused about the appropriateness of the numbers in the simulations. In the new Figure S9, it appears that the expression of vimentin can range over 20 orders of magnitude - surely this is nonbiological, nonbiophysical, and nonrealistic? Even the nine orders of magnitude range (10^-4 - 10^5) mentioned in the response to reviewers document (possibly this number is in reference to cadherin?) seems unfeasibly large. If the model produces numbers that are so at odds with the authors' own experimental data (e.g. flow cytometry data), we cannot have confidence in its predictions and potential biological insights.”*

We thank the editor for highlighting this important concern. However, we respectively (but strongly) disagree with the conclusion that our model is infeasible. A major requirement of computational models is the identification and validation of the model by replicating experimental datasets. As the editor knows, in signal transduction modeling, these data sets are often qualitative, for example immunoblots such as Western blots. As is typically true for many (if not most) signal transduction models found in the literature, in this study we used immunoblots to both estimate and validate model parameter sets. However, because we used primarily Western blot information, the model concentration scale is arbitrary. Toward this issue, we have previously published the POETs technique which estimates model parameters by comparing both scaled states and an order of magnitude estimate of the physiological concentration scale. In this case, we used pM as our “natural scale”. Thus, when we have values of 1000 - 10000 for a simulated state we are in the nM range (which is physiologically reasonable). On the low end of the scale, POETs does not constrain for “zero values” thus, small values indicate no expression. Given that our natural scale is pM, we use a cutoff of 10^-3 to indicate no expression. Thus, below 10^-3 states are practically indistinguishable from zero. We have updated the model estimation section of the materials and methods, as well as added additional text to the expanded discussion of POETs presented in the supplemental data to further clarify this important point.

In the case of our studies, Figure 2 identifies experimental data extracted from published Western blots and our simulation results. It is clear from these that 1) the training data included a variety of Western blot data treatments that were effectively zero, and 2) our simulations matched the training data virtually perfectly over time and across multiple biological species. These results validate the power of our simulation scheme, which necessarily includes numerical interpretations of the no expression state. In our Supplemental Figure S9, we presented the raw concentrations of our simulations (previously validated from the training data) over hundreds of parameter ensembles, *with no “zero” threshold applied*. In these results, we identify expressions of E-cadherin between 10-6 – 104 and vimentin from 10-15 to 105. While we present the absolute computational values, we only interpret the data as any value below 10-3 equivalent to zero. Indeed, our robustness coefficients (Figure 3), **which represent the ratio of integrated concentration of the treatment effect over the baseline effect,** identify no difference in model output for any vimentin (or e-cadherin for that matter) value less than 10-3, supporting this interpretation. The very large robustness coefficients for Vimentin are a numerical artifact of the treated/baseline cases (divided by the no expression case). Furthermore, the actual numerical value for a robustness coefficient is not important; these are qualitative measures similar to sensitivity. Rather, it the qualitative shift following treatment that is important. Lastly, the robustness analysis identified a subpopulation that didn’t behave as expected. We used this qualitative result to design a series of experiments that not only found this subpopulation, but also validated the specific signaling pathways used in this subpopulation. Thus, we feel very strongly that robustness analysis is a valuable qualitative approach that was directly experimentally validated in our study.

To reiterate, any interpretation of our data at extremely small (<10-3) or extremely large (<106) values was taken into account, and not used for biological comparison.

To clarify, we have revised figures S9 and 3:

1. Figure S9 – We have shaded the areas which are not biological relevant and moved the origin accordingly (anything less than 10E-3). This should help the readers focus only on the relevant data points.
2. Figure 3 – Region P3 now includes all values from 10E-3 to 10E-15. Because these values are “biological zero” all dots within this range behave the same (no response to stimuli). All other regions remain the same.

Robustness analysis is a useful tool to identify the relative integrated response (area under the curve for the treated case/area under the curve baseline) of parameter sets in the parameter ensembles. Robustness coefficients allow us to find parameter sets which: 1) give the same qualitative outcome of a biological model, 2) can identify whether similar or different paths through the biological network, and 3) can identify critical and dispensable nodes. Thus, the absolute magnitude of robustness coefficients is not important. Rather, the shift in coefficients gives information about changes in biological function following stimulation. This is similar to sensitivity coefficients, which are widely used in the field (and can vary over many orders of magnitude).

Robustness analysis has been previously used in a number of published studies to identify biological network performance information that was inaccessible with other approaches (Lequieu, PLoS Comput Biol. 2011 Nov; 7(11): e1002263; Tasseff, Integr Biol (Camb) 2011;3:578–591; Tasseff, PloS One, DOI: 10.1371/journal.pone.0008864). Given the fact that robustness is a ratio of integrated concentration curves (comparing stimulated vs. unstimulated conditions), one could expect a wide range of values (similar to what is seen with a sensitivity analysis), especially if the unstimulated conditions are close to zero (<10E-3 and dividing by this number) or if the concentration values grow exponentially. This tool has been extremely successful in identifying relevant signaling pathways and critical nodes as previously published. As previously mentioned, the scaling of values and thresholds are commonly employed for sensitivity/robustness analyses, which we neglected to do here so that others could more easily replicate our results.

Again, any interpretation of our data at extremely small (<10-3) or extremely large (<106) values was taken into account, and not used data interpretation for biological comparison.

With regard to the flow cytometry data, we believe this actually supports our case because it can track expression in individual cells. If we had 1 million cells to scan, we technically have 6 orders of possible expression in the population (from 0 to all cells express), or 9 orders with a billion cells. Furthermore, the numerical effects of absolute zero are such that log scales are employed that have 10^-x to approach zero. Similar to creating a “zero” threshold in computation, gating is widely used to analyze relevant data points within a specific region. We do not consider "negative" expression values, as the editor and we rightly agree wouldn't make sense, but fractional expression (0<x<10E6) is computationally reasonable. We therefore believe our regional plots are biological meaningful, and any interpretation/prediction of the data was only used for the relevant regions.

Lastly, we are confident that if any reader or collaborator uses our model to study TGFb/VEGF regulated EMT, they will get similar results for the same inputs. Numerous times we have tried to utilize previously published systems models of varying sizes that contained poorly or non-articulated normalization or scaling schemes and are unable to acquire the same output. We thought about applying a similar scheme herein, and argue that not having done so apriori has no bearing on the quality of the data or the interpretation. Indeed, Reviewer 1 also recognized the significance of our modeling strategy and its particular formatting so that others can replicate and use it. The other reviewers also had no concerns with this approach or the numbers we are reporting.

**Reviewer 1**

*I appreciate that the authors addressed a very important biological problem, and their efforts to share their model with the community so others can reproduce their results. The authors have done extensive experimental studies, which are well performed. Their study on the TGF and VEGF-A pathway crosstalks is timely.*

We appreciate your enthusiasm for our model, its timeliness, and how we share this model with the community so that it can be easily replicated without loss in interpretation.

*“I am concerned about the modeling efforts. The story can be very impressive and highly significant to the community, but is severely affected by the extremely unnecessary complexity of modeling efforts. My concern is on a fundamental question: why do we do modeling studies? There are many ways and approaches to model a biological system, depending on the problem and on the availability of information. Despite all these, we do modeling studies because they can tell us something new, give us new insight that we may otherwise not easy to get. There are a number of modeling studies on EMT. Lu et al. (Lu et al. 2013) and Tian et al. (Tian, Zhang, and Xing 2013) both built nice and simple models of EMT regulation. Both of these studies suggest clear molecular mechanisms, and make testable predictions, which lead to a following experimental test. The work of Steinway et al (Steinway et al. 2014) is also on pathway crosstalks in EMT. These researchers also worked on a rather large network. They did some coarse-graining to focus on the key network structure, and from it they did some analysis to help understand how pathway crosstalks may affect EMT dynamics. On the other hand, if for a network that is so complex we can’t even examine it visually (not even through modular analysis), how can we know the validity of the model and gain new insight? We know that every model (in biology) is incomplete, but we need to know where we have made assumptions, and where we may miss components. The authors acknowledge that the network does not include miRNAs since the information is incomplete---very fair statement, but then how should a reader trust the work without this important component? You may argue that this criticism applies to every model. The point is that for a “simpler” model one knows what might be missing and what might be the consequences, also what might have been treated implicitly in the model.”*

This comment reads more like a philosophical muse about the utility of modeling rather than a direct criticism of the work itself, and indeed the reviewer suggests as much. The reviewer expresses criteria that justify the use of a modeling approach. Such useful simulations 1) tell us something new, 2) predict emergent phenomena that would not have been determinable via experiment alone, and 3) have predictions that be validated by experimental tests motivated by the system. The reviewer then cites studies about EMT that they believe satisfy these criteria. We fully concur that the models cited by the reviewer are simpler than our model presented here, and that they too identify and validate testable predictions about the interactions contained. However, we are fully confident that our model similarly satisfies these criteria as demonstrated by the data presented in this manuscript. Our computational model 1) identified the existence of a hybrid EMT state which was not part of the published literature used for training of the model, 2) predicted that EMT was controlled primarily by pSP1 and NFAT, and 3) this prediction was experimentally proven as was the predicted expression required to create the hybrid EMT state. In many ways, computational simulations work within a “control volume” of all the components that are simulated, thus any emergent behaviors identified arise out of combinations of the components. For very small models, the possible outcomes are few but with strong potential for non-linear relationships to be predicted. For very large models, the possible solution states are tremendous, with less of an ability to predict nonlinearity between any node pair. The utility of each model depends on the question you want to ask. For a small model, one is primarily interested in how a few components could interact to create a mechanism that is bounded by those components. For a large model, the objective is the same, but a major difference is that the possible experimental space is truly intractable without the model. In each case, there are unknown interactions from outside of the control volume that could affect the overall solution. If these were dominating however, each model’s prediction would not be validated by the applied experiments. Neither in the papers cited by the reviewer, nor in our manuscript, was this the case – thus each model were highly useful for their purpose and outcome. We argue here that the large model approach was essential to the results and impact we report. Our simulations integrate two very different ligand inputs through a huge (yet physiologically valid) network of potential interacting species) and end on two very different protein expression outputs. We agree that there should be no reason to expect specific testable outcomes with our model if it were needlessly complex and spurious – unless it in fact contains sufficient information. The only way to know is if the simulation creates predictions of emergent phenomena that are experimentally testable. Our model does so as demonstrated by the data presented. The congruency between the data and the model the reviewer does not dispute. His/her concern with the lack of consideration of miRNAs (a valid limitation that is so mentioned) is that our results may be somehow compromised as to give a uniquely validated prediction through false means. Considering the biological hierarchy, miRNAs are between the DNA and RNA. In our modeling effort we have modeled several components like DNA, RNA, proteins, protein complexes and so on. For modeling the system, at the end it boils down to which components can effect, how do they effect and how can we capture their effect. While, we have not explicitly modeled miRNA components, the effect of this missing component as the reviewer points out is potentially incorporated in the values of the different family of parameter sets that we have obtained. So irrespective of the fact whether a particular miRNA component had a role or not, using our model we could satisfies multiple experimental datasets across time and ligand dose from multiple published reports. Indeed, with the incorporation of miRNA networks (once they are known) and anything else that is discovered, we may further understand how this larger system behaves and identify more phenomena. However any subsequent analysis and experimentation will still have to satisfy these conditions and results as part of its validation process.

*I hope that the authors do not take my words as against the work. Instead I think that they work on an important problem and can make a nice story. Just try to make the story as simple as possible: if one can use a coarse-grained model (instead of mass-action type) to show how the pathways crosstalk and generate some measurable effects which are supported by their data, why use such a complicated model with unbelievable number of parameters that most researchers would cast doubt on it (and on the work as a whole)—the number of parameters make all the model-experimental agreements not as impressive as it should be. Also a simpler model can guide them to understand better of the system. For example, they observed the hybrid state. Can they say something on the mechanism like the papers I mentioned above? Observing the hybrid state alone is not surprising given the many feedback loops in the system.*

*As already stated in our manuscript, we have a lengthy discussion paragraph on why we chose our large scale modeling approach, and how others have used similar approaches to identify biological relevance. We have highlighted this section again in the manuscript (lines 495-535). Just a few sentences to re-iterate our point:*

***“……………We used the traditional approach of mass action kinetics within an ordinary differential equation framework. The identification problem for the EMT model was massively underdetermined. This is not uncommon for differential equation models, especially those that are highly mechanistic. Of course, we could have discarded mechanism or reduced the model scope to decrease the complexity of the identification problem. However, a central criticism leveled by biologists is that model simplification is often done at the cost of biological reality, or done for reasons of computational expediency (Sainani et al., 2012). To avoid this criticism, we systematically identified an ensemble of likely models each consistent with the training data, instead of a single but uncertain best fit model. Previously, we (and others) have suggested that deterministic ensembles could model heterogeneous populations in situations where stochastic computation was not feasible (Lequieu, et al. 2011). Population heterogeneity using deterministic model families has previously been explored for bacterial growth in batch cultures (Lee et al., 2011)……………..”***

The reviewer is essentially making our point by pointing out criteria that our simulations and experimental data satisfy. We argue the phenomena and mechanisms predictable from large models are far broader than in isolated small models (though not necessarily more important). Therefore each modeling approach has its benefits. Indeed, the results of our simulations and experimentation can be further used to reduce our model into a potentially coarser grained network, but this could not have been done apriori. Furthermore, other investigators can use our validated simulations to both drill down into subsets of the model and test other signaling combinations we did not. Therefore, this large model creates more opportunities for scientific discovery across a broader space of science.

Regarding the existence of feedback loops and large parameter spaces in large models, our previously published POETs strategy enables robust predictions of specific outcomes with confined parameter sets. As demonstrated by our simulations and data, there is not only one outcome possible but multiple, yet these outcomes are highly regulated by two principal molecular components that could not be determined apriori. If these simulations were as spurious as the reviewer poses, then multiple signaling states would achieve the same outcome, or conversely the same signaling states would achieve different outcomes. We conclusively determine neither occurs, thus substantiating the mechanism for the phenomena we report.

We may not be able to completely assuage the reviewer about the value of large models, but it is clear that the readership of this journal very much appreciates the power of such simulations, including those using mass-action kinetics, as demonstrated by a cursory analysis of the literature.

<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0030189>(cited by 329)

<http://www.ncbi.nlm.nih.gov/pubmed/19156131> (cited by 280)